UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

A New Computational Method for the Sparsest Solutions to Systems of Linear Equations

Zhao, Yun-Bin; Kocvara, Michal

DOI:
10.1137/140968240

Document Version
Peer reviewed version

Citation for published version (Harvard):
Zhao, Y-B & Kocvara, M 2015, 'A New Computational Method for the Sparsest Solutions to Systems of Linear Equations', *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 1110-1134. https://doi.org/10.1137/140968240

Link to publication on Research at Birmingham portal

Publisher Rights Statement:
© 2015, Society for Industrial and Applied Mathematics
Read More: http://epubs.siam.org/doi/10.1137/140968240

Eligibility for repository checked July 2015

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 05. May. 2023

A NEW COMPUTATIONAL METHOD FOR THE SPARSEST SOLUTIONS TO SYSTEMS OF LINEAR EQUATIONS

YUN-BIN ZHAO* AND MICHAL KOČVARA†

Abstract. The connection between the sparsest solution to an underdetermined system of linear equations and the weighted ℓ_1-minimization problem is established in this paper. We show that seeking the sparsest solution to a linear system can be transformed to searching for the densest slack variable of the dual problem of weighted ℓ_1-minimization with all possible choices of nonnegative weights. Motivated by this fact, a new reweighted ℓ_1-algorithm for the sparsest solutions of linear systems, going beyond the framework of existing sparsity-seeking methods, is proposed in this paper. Unlike existing reweighted ℓ_1-methods that are based on the weights defined directly in terms of iterates, the new algorithm computes a weight in dual space via certain convex optimization and uses such a weight to locate the sparsest solutions. It turns out that the new algorithm converges to the sparsest solutions of linear systems under some mild conditions that do not require the uniqueness of the sparsest solutions. Empirical results demonstrate that this new computational method remarkably outperforms ℓ_1-minimization and stands as one of the very efficient sparsity-seeking algorithms for the sparsest solutions of systems of linear equations.

Key words. ℓ_0-minimization, sparsest solution, reweighted ℓ_1-method, convex optimization, linear programming, bilevel optimization, sparsity recovery

AMS subject classifications. 90C25, 90C26, 15A06, 90C05, 15A29, 65K05

1. Introduction.

Sparsity has long been utilized in the signal and image processing community (see, e.g., Beurling [5], Pennebaker and Mitchell [49], Gorodnitsky, George, and Rao [32], Donoho [23], and Mallat [42]) and dates back to the early 1900's (e.g., Carathéodory [13]). It has long been exploited in learning theory and statistics as well (e.g., Tibshirani [51], Mangasarian [44], Vapnik [55], and Hastie, Tibshirani, and Friedman [34]). The compressed sensing, initiated by Candès, Romberg, and Tao [11, 9, 10] and Donoho [24], has attracted considerable cross-disciplinary attention in recent years and stimulates a plethora of new applications of sparsity in such fields as geophysical data analysis, medical imaging, communications, sensor network, and computational biology. A central question in these applications can be cast, typically, as the so-called ℓ_0-minimization problem

$$\min\{\|x\|_0 : Ax = b\}, \tag{1.1}$$

where $A \in R^{m\times n}$ $(m < n)$ is a given full-rank matrix (i.e., $\mathrm{rank}(A) = m$), $b \in R^m$ is a given vector, and $\|\cdot\|_0$ counts the nonzeroes of a vector. We assume $b \neq 0$ throughout the paper. Clearly, the optimal solution of (1.1) is the sparsest solution to the linear system $Ax = b$.

Developing efficient algorithms to solve the problem (1.1) is fundamentally important and has become a common request in various applications. Over the past few years, some practical methods have been proposed for this problem, including the greedy pursuit (see, e.g., [43, 22, 52, 25, 6, 19, 48, 7]) and convex optimization (e.g.,

*School of Mathematics, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom (y.zhao.2@bham.ac.uk). The research of this author was supported by the Engineering and Physical Sciences Research Council (EPSRC) under grant #EP/K00946X/1.

†School of Mathematics, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom (m.kocvara@bham.ac.uk), and Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 18208 Praha 8, Czech Republic. The research of this author was partly supported by the Grant Agency of the Czech Republic through project GAP201-12-0671.

[16, 27, 31, 52, 11, 10, 53]). Closely related to greedy pursuits are thresholding methods (see, e.g., [20, 6, 47, 4, 41]), which also attract considerable recent attention in the field of sparse signal recovery. Other methods, such as nonconvex optimization and Bayesian framework, are exploited by some researchers as well (e.g., [14, 56]). A good introduction and survey of these methods can be found in [8, 28, 54]. Particularly, ℓ_1- and weighted ℓ_1-minimization play a vital role in the development of compressed sensing theory, and have been widely used for solving the ℓ_0-problem (1.1).

The efficiency of ℓ_1-minimization has been analyzed (largely in the context of compressed sensing) under various assumptions such as the mutual coherence [26, 28], restricted isometry property (RIP) [11], null space property (NSP) [18], exact recovery condition (ERC) [52, 31], and the range space property [58]. These analyses also motivate ones to explore a more efficient method than ℓ_1-minimization. Candès, Wakin, and Boyd [12] have proposed the reweighted ℓ_1-minimization and have empirically demonstrated that this method often outperforms ℓ_1-minimization. Needell [46] has analyzed this method in the noisy case and obtained an improved stability result under the RIP assumption. Asif and Romberg [1] have presented a homotopy method to solve the reweighted ℓ_1-minimization inexpensively as the weight changes. They also utilized the reweighted ℓ_1-method to cope with the sparse recovery of streaming signals (from streaming measurements) (see [2]). Zhao and Li [59] have shown that a large family of reweighted ℓ_1-algorithms converges to sparse solutions of underdetermined linear systems under certain assumptions imposed on the matrix. This family includes many existing reweighted ℓ_1-algorithms (e.g., [12, 30, 39, 57, 17]) as special cases. Moreover, the reweighted ℓ_1-method is also used for the study of partial-support-information-based signal recovery (see Khajehnejad et al. [38]).

Close to reweighted ℓ_1-algorithms is the reweighted ℓ_2-method which has a relatively long history (see, e.g., [35, 32, 29, 15, 57, 21, 3, 40, 50]). As pointed out by Wipf and Nagarajan [57], both reweighted (ℓ_1- and ℓ_2-) methods can be derived by estimating the upper bound of certain sparsity merit functions. It is worth noting that the main difference between various reweighted ℓ_1-methods lies in the updating scheme of weights (see, e.g., [30, 39, 57, 37, 57, 17, 36, 59]). A common feature for these methods is that the magnitude of the weight w^k is determined locally at the current iterate x^k. The weight is often chosen to penalize those components of the solution, which correspond to the small components of the current iterate. Such a weight might force the next iterate to admit a sparsity pattern very similar to the current iterate, and thus the next iterate may still fail to change toward the sparsest pattern of solutions if the current iterate is far from being the sparsest.

In this paper, we develop a new reweighted ℓ_1-method to locate the sparsest solution of linear systems. A unique feature of this method is that the weight is computed in certain dual space via convex optimization, instead of being defined locally on the current iterate. We first employ the strict complementarity theory of linear programs to prove that ℓ_0-minimization can be reformulated as an equivalent ℓ_0-maximization problem with bilevel constraints. As a result, solving ℓ_0-minimization can be translated to the computation of an optimal weight, which can be achieved, in theory, by searching the densest possible slack variable of the dual problem of weighted ℓ_1-minimization with all possible weights. This connection between the sparsity in primal space and density in dual (complementary) space leads to a new computational method for solving ℓ_0-minimization, which goes beyond the framework of existing sparsity-seeking methods and does not share the aforementioned common feature. Under certain conditions, we prove that the proposed algorithm converges to

the sparsest solutions of systems of linear equations. Our convergence analysis permits the linear system to admit multiple sparsest solutions. The perturbation theory for linear programs, established by Mangasarian and Meyer [45], plays a vital role in our analysis. Empirical results indicate that this computational method remarkably outperforms the standard ℓ_1-minimization, and its performance is very comparable to the state-of-art reweighted ℓ_1-algorithms for ℓ_0-problems.

In section 2, we establish some theoretical results on the relationship between ℓ_0- and weighted ℓ_1-minimization and provide an intrinsic connection between bilevel optimization and ℓ_0-minimization. Based on the results in section 2, we propose a new reweighted ℓ_1-method in section 3. The convergence analysis for this method is carried out in section 4, and the numerical results are given in section 5.

Notation. In this paper, all vectors are column vectors, unless stated otherwise. R^n_+ (R^n_{++}) is the set of nonnegative (positive) vectors in R^n. We interchangeably use $w \in R^n_+$ ($w \in R^n_{++}$) and $w \geq 0$ ($w > 0$), and we use e to denote the vector of ones, i.e., $e = (1, 1, ..., 1)^T$. For a given subset $J \subseteq \{1, 2, ..., n\}$ and a matrix $A \in R^{m\times n}$, A_J denotes the submatrix of A consisting of the columns indexed by J, and A^T_J is the transpose of A_J. Similarly, for the vector $x \in R^n$, x_J denotes the subvector of x indexed by J, and x^T_J is the transpose of x_J, and we denote by $J_+(x) = \{i : x_i > 0\}$, $J_-(x) = \{i : x_i < 0\}$ and $J_0(x) = \{i : x_i = 0\}$. Clearly, $J_+(x) \cup J_-(x) = supp(x) := \{i : x_i \neq 0\}$, the support of x. $\mathcal{R}(A^T) = \{A^T y : y \in R^m\}$ denotes the range space of A^T. For $x, y \in R^n$, the inequality $x \leq y$ ($x < y$) means $x_i \leq y_i$ ($x_i < y_i$) for all $i = 1, ..., n$. We use $|\cdot|$ throughout the paper: For a set S, $|S|$ denotes its cardinality; for a vector $x = (x_1, ..., x_n)^T$, $|x|$ is the absolute value of x, i.e., $|x| = (|x_1|, ..., |x_n|)^T$; for a matrix $A = (a_{ij})$, $|A|$ stands for the absolute version of A, i.e., $|A| = (|a_{ij}|)$.

2. Sparsity and density. In this section, we develop some theoretical results concerning the connection between ℓ_0- and weighted ℓ_1-minimization. These results provide an incentive to develop a new computational method for ℓ_0-problems (see section 3 for details). Let us first recall the following result.

Lemma 2.1 (Theorem 2.10 in [58]). *x is the unique solution to the ℓ_1-problem $\min\{\|x\|_1 : Ax = b\}$ if and only if the following conditions hold: $(A_{J_+(x)}, A_{J_-(x)})$ has a full-column rank, and there exists a vector $\eta \in \mathcal{R}(A^T)$ such that $\eta_{J_+(x)} = e_{J_+(x)}$, $\eta_{J_-(x)} = -e_{J_-(x)}$, and $\|\eta_{J_0(x)}\|_\infty < 1$.*

The above sufficiency and necessity were established in [31] and [58], respectively. By Lemma 2.1, we can characterize the uniqueness of solutions to weighted ℓ_1-minimization.

Theorem 2.2. *Let $w \in R^n_{++}$ be a given weight and $W = diag(w)$. Then x is the unique solution to the weighted ℓ_1-problem*

$$\min\{\|Wx\|_1 : Ax = b\} \tag{2.1}$$

if and only if the following conditions hold: $(A_{J_+(x)}, A_{J_-(x)})$ has a full-column rank, and there is a vector $\xi \in \mathcal{R}(A^T)$ satisfying that $\xi_{J_+(x)} = w_{J_+(x)}$, $\xi_{J_-(x)} = -w_{J_-(x)}$ and $|\xi_{J_0(x)}| < w_{J_0(x)}$.

Proof. By setting $u = Wx$, the problem (2.1) can be written as

$$\min\{\|u\|_1 : (AW^{-1})u = b\}. \tag{2.2}$$

Since W is a diagonal matrix with positive diagonal entries, we see that u and x have the same support sets and $J_+(u) = J_+(x)$ and $J_-(u) = J_-(x)$. Clearly, x is the unique

solution to the problem (2.1) if and only if u is the unique solution to the problem (2.2). By Lemma 2.1, u is the unique solution to (2.2) if and only if the following two conditions hold: (i) $((AW^{-1})_{J_+(u)}, (AW^{-1})_{J_-(u)})$ has a full-column rank; (ii) there exists a vector $\eta \in \mathcal{R}((AW^{-1})^T)$ such that

$$\eta_{J_+(u)} = e_{J_+(u)},\ \eta_{J_-(u)} = -e_{J_-(u)},\ \|\eta_{J_0(u)}\|_\infty < 1. \tag{2.3}$$

Since $J_+(u) = J_+(x), J_-(u) = J_-(x)$ and W^{-1} is a diagonal nonsingular matrix, we deduce that the condition (i) above is equivalent to that $(A_{J_+(x)}, A_{J_-(x)})$ has a full-column rank. Note that $\eta \in \mathcal{R}((AW^{-1})^T)$ is equivalent to $\xi = W\eta \in \mathcal{R}(A^T)$. The condition (2.3) is equivalent to $\xi_{J_+(x)} = w_{J_+(x)},\ \xi_{J_-(x)} = -w_{J_-(x)}$ and $|\xi_{J_0(x)}| = |(W\eta)_{J_0(x)}| < w_{J_0(x)}$. □

For any x^* satisfying $Ax^* = b$, if $A_{supp(x^*)}$ has a full-column rank, we can prove that there is a weight w such that x^* is the unique solution to the problem (2.1).

Theorem 2.3. *Let x^* be a solution to the system $Ax = b$ with $A_{supp(x^*)}$ having a full-column rank. Let $w \in R^n_{++}$ with $w_{J_+(x^*)}$ and $w_{J_-(x^*)}$ being given. If $w_{J_0(x^*)} > \eta^*$ where η^* is an optimal solution to the linear program*

$$\min_{(y,\eta)}\{e^T\eta : A^T_{J_+(x^*)}y = w_{J_+(x^*)},\ A^T_{J_-(x^*)}y = -w_{J_-(x^*)},\ -\eta \le A^T_{J_0(x^*)}y \le \eta\}, \tag{2.4}$$

then x^* is the unique solution to the weighted ℓ_1-problem $\min\{\|Wx\|_1 : Ax = b\}$, *where $W = \mathrm{diag}(w)$.*

Proof. When $(A_{J_+(x^*)}, A_{J_-(x^*)})$ has a full-column rank, the range space of $(A_{J_+(x^*)}, A_{J_-(x^*)})^T$ is the whole space $R^{|J_+(x^*)|+|J_-(x^*)|} = R^{|supp(x^*)|}$. Thus for any vector $u \in R^{|J_+(x^*)|}$ and $v \in R^{|J_-(x^*)|}$, there always exists a vector $y \in R^m$ such that $A^T_{J_+(x^*)}y = u$ and $A^T_{J_-(x^*)}y = -v$. Hence, the problem (2.4) is always feasible for any given $(w_{J_+(x^*)}, w_{J_-(x^*)}) > 0$. Let (y^*, η^*) be an optimal solution to the problem (2.4). Clearly, we must have that $\eta^* = |A^T_{J_0(x^*)}y^*|$. Let $w_{J_0(x^*)} > \eta^*$. By setting $\xi = A^Ty^*$, we see from the problem (2.4) that

$$\xi_{J_+(x^*)} = w_{J_+(x^*)},\ \xi_{J_-(x^*)} = -w_{J_-(x^*)},\ |\xi_{J_0(x^*)}| = |A^T_{J_0(x^*)}y^*| = \eta^* < w_{J_0(x^*)}.$$

Thus, by Theorem 2.2, x^* is the unique solution to the weighted ℓ_1-problem. □

In Theorems 2.2 and 2.3, the weight w is required to be positive. However, this is not required in the next result.

Theorem 2.4. *Let x^* be a solution to the system $Ax = b$ with $A_{supp(x^*)}$ having a full-column rank. If $w \in R^n_+$ satisfies that*

$$w_{J_0(x^*)} > \left|A^T_{J_0(x^*)}A_{supp(x^*)}(A^T_{supp(x^*)}A_{supp(x^*)})^{-1}\right| w_{supp(x^*)}, \tag{2.5}$$

then x^* is the unique solution to the weighted ℓ_1-problem $\min\{\|Wx\|_1 : Ax = b\}$, *where $W = diag(w)$.*

Proof. Let x^* be a solution to the linear system $Ax = b$ and let $A_{supp(x^*)}$ have a full-column rank. Denote by $J = supp(x^*)$ and $J_0 = J_0(x^*)$ for simplicity. Let y be an arbitrary solution to the linear system. Since $A_Jx^*_J = b$ and $A_Jy_J + A_{J_0}y_{J_0} = b$, we have $A_J(y_J - x^*_J) + A_{J_0}y_{J_0} = 0$. Thus

$$x^*_J = y_J + (A^T_JA_J)^{-1}A^T_JA_{J_0}y_{J_0}, \tag{2.6}$$

which implies that $|x^*_J| \le |y_J| + |(A_J^TA_J)^{-1}A_J^TA_{J_0}| \cdot |y_{J_0}|$. Therefore,

$$\begin{aligned}\|Wx^*\|_1 - \|Wy\|_1 &= w^T|x^*| - w^T|y| = w_J^T|x^*_J| - w_J^T|y_J| - w_{J_0}^T|y_{J_0}| \\ &\le w_J^T|(A_J^TA_J)^{-1}A_J^TA_{J_0}| \cdot |y_{J_0}| - w_{J_0}^T|y_{J_0}| \\ &= \left(|A_{J_0}^TA_J(A_J^TA_J)^{-1}|w_J - w_{J_0}\right)^T |y_{J_0}|.\end{aligned}$$

For any solution $y \neq x^*$, we see from (2.6) that $y_{J_0} \neq 0$. From (2.5) and the inequality above, we infer that $\|Wx^*\|_1 - \|Wy\|_1 < 0$ for any solution $y \neq x^*$ to the system $Ax = b$. Thus x^* is the unique solution to the weighted ℓ_1-problem. $\square$

While a certain relationship between Theorems 2.3 and 2.4 exists, these results are different in general. In (2.5), the components $w_i, i \in supp(x^*)$, are allowed to be zero, while these components are positive in (2.4). It is well known that at any sparsest solution x^* of the system $Ax = b$, the associated matrix $A_{supp(x^*)}$ always has a full-column rank. Thus the following corollary is an immediate consequence of Theorems 2.3 and 2.4.

Corollary 2.5. *Let x^* be a sparsest solution to the system $Ax = b$. Let $W = diag(w)$, where w satisfies one of the following conditions:*

(i) $w \in R^n_+$ *and* $w_{J_0(x^*)} > \left|A^T_{J_0(x^*)}A_{supp(x^*)}(A^T_{supp(x^*)}A_{supp(x^*)})^{-1}\right| w_{supp(x^*)}$.

(ii) $w \in R^n_{++}$ *and* $w_{J_0(x^*)} > \eta^*$, *where η^* is an optimal solution of (2.4).*

Then x^* is the unique solution to the weighted ℓ_1-problem $\min\{\|Wx\|_1 : Ax = b\}$.

Thus for any sparsest solution x^*, there exists a weight accordingly such that x^* is the unique optimal solution to the weighted ℓ_1-problem. Clearly, such a weight (as indicated by Corollary 2.5) is not unique. Although the choice of such weights depends on the support of the sparsest solution, Corollary 2.5 remains very useful for the development of some theoretical properties and computational methods for ℓ_0-problems. Indeed, based on Corollary 2.5 and the strict complementarity theory of linear programs, we can reformulate ℓ_0-minimization as a structured bilevel optimization problem, which eventually leads to a practical algorithm for ℓ_0-minimization. Note that the problem (2.1) can be written as

$$\min_{(x,t)}\{w^Tt : \ Ax = b, \ |x| \le t\}. \tag{2.7}$$

Clearly, x^* is an optimal solution to the problem (2.1) if and only if (x^*, t^*), where $t^* = |x^*|$, is an optimal solution to the problem (2.7). By introducing variables $\alpha, \beta \in R^n_+$, the problem (2.7) can be further written as

$$\min_{(t,x,\alpha,\beta)}\{w^Tt : \ t - x - \alpha = 0, \ -t - x + \beta = 0, \ Ax = b, \ (t, \ \alpha \ \beta) \ge 0\}. \tag{2.8}$$

Let $(x^*, t^*, \alpha^*, \beta^*)$ be an optimal solution to the problem above. Clearly, $t^* = |x^*|$, $\alpha^* = t^* - x^* = |x^*| - x^*$, and $\beta^* = t^* + x^* = |x^*| + x^*$. The dual problem of (2.8) is given as

$$\max_{(u,z,y)}\{b^Ty : \ u - z \le w, \ u + z - A^Ty = 0, \ -u \le 0, \ z \le 0\}.$$

By setting $v = -z$ and $s = w - (u - z) = w - u - v$, the problem can be written as

$$\max_{(s,y,u,v)}\left\{b^Ty : \ A^Ty - u + v = 0, \ s = w - u - v, \ (s, u, v) \ge 0\right\}. \tag{2.9}$$

It should be noted that some of variables in (2.9) can be eliminated to slightly simplify the problem and the statement of the algorithm in later sections. However, we retain these variables since it is convenient to include them as we carry out the convergence analysis in section 5. By the complementarity property, the optimal solutions of (2.8) and (2.9) satisfy the complementarity conditions

$$t^T s = 0,\ \alpha^T u = 0,\ \beta^T v = 0,\ (t, \alpha, \beta, s, u, v) \geq 0, \tag{2.10}$$

and hence $\|t\|_0 + \|s\|_0 \leq n$. So the sparsity of t in the original problem (2.8) is closely related to the density of the variable s in the dual problem (2.9). Moreover, by linear programming theory, there exists a pair of solutions to (2.8) and (2.9) which are strictly complementary in the sense that they satisfy (2.10) and $t + s > 0$, $\alpha + u > 0$ and $\beta + v > 0$. We summarize this fact as follows.

Lemma 2.6. *Let $w \in R^n_+$ be given. Then x^* is an optimal solution to the problem (2.1) if and only if $(x^*, t^*, \alpha^*, \beta^*)$ is an optimal solution to the problem (2.8). For any optimal solution $(x^*, t^*, \alpha^*, \beta^*)$ of (2.8), we must have that $t^* = |x^*|$, $\alpha^* = |x^*| - x^*$ and $\beta^* = |x^*| + x^*$. Moreover, there always exists a solution $(t^*, x^*, \alpha^*, \beta^*)$ to (2.8) and a solution (s^*, y^*, u^*, v^*) to (2.9) such that t^* and s^* are strictly complementary, i.e., $(t^*, s^*) \geq 0, (t^*)^T s^* = 0$ and $t^* + s^* > 0$.*

From the above lemma, we must have $t^* = |x^*|$ (where x^* is the optimal solution to the weighted ℓ_1-problem (2.1)). The strict complementarity of s^* and t^* implies that $\|x^*\|_0 + \|s^*\|_0 = \|t^*\|_0 + \|s^*\|_0 = n$. Thus, if s^* is the densest variable of (2.9), then x^* must be the sparsest solution of the linear system. We now prove that a certain bilevel optimization problem (which seeks the density of the dual variable s) provides an optimal weight w^*, by which the weighted ℓ_1-minimization yields the sparsest solution of the linear system.

Theorem 2.7. *Let $(s^*, y^*, w^*, u^*, v^*, \gamma^*)$ be an optimal solution to the bilevel optimization*

$$\begin{aligned} \max_{(s,y,w,u,v,\gamma)} \ & \|s\|_0 \\ s.t. \quad & b^T y = \gamma,\ A^T y - u + v = 0,\ s = w - u - v,\ (s, u, v) \geq 0, \\ & w \geq 0,\ \gamma = \min_x \{\|Wx\|_1 : Ax = b\}, \end{aligned} \tag{2.11}$$

where $W = \mathrm{diag}(w)$. Then any optimal solution to the weighted ℓ_1-problem (2.1) with $w = w^*$ is a sparsest solution to the system $Ax = b$.

Proof. Let $\widehat{x}$ be a sparsest solution to the system $Ax = b$. By Corollary 2.5, there exists a weight $\widehat{w} \in R^n_+$ such that $\widehat{x}$ is the unique solution to the weighted ℓ_1-problem $\min\{\|\widehat{W}x\|_1 : Ax = b\}$, where $\widehat{W} = \mathrm{diag}(\widehat{w})$. Then by Lemma 2.6, $(\widehat{t}, \widehat{x}, \widehat{\alpha}, \widehat{\beta})$ is the unique solution to the problem (2.8), where $\widehat{t} = |\widehat{x}|$, $\widehat{\alpha} = |\widehat{x}| - \widehat{x}$ and $\widehat{\beta} = |\widehat{x}| + \widehat{x}$. Consider the dual problem of the above weighted ℓ_1-minimization

$$\max_{(s,y,u,v)} \{b^T y :\ A^T y - u + v = 0,\ s = \widehat{w} - u - v,\ (s, u, v) \geq 0\}.$$

By Lemma 2.6, there exists an optimal solution to the above dual problem, denoted by $(\widehat{s}, \widehat{y}, \widehat{u}, \widehat{v})$, such that $(\widehat{t}, \widehat{\alpha}, \widehat{\beta})$ and $(\widehat{s}, \widehat{u}, \widehat{v})$ are strictly complementary. In particular, $\widehat{t}$ and $\widehat{s}$ are strictly complementary. Thus $\|\widehat{t}\|_0 + \|\widehat{s}\|_0 = n$ and

$$\|\widehat{x}\|_0 = \|\widehat{t}\|_0 = n - \|\widehat{s}\|_0. \tag{2.12}$$

By linear programming strong duality, we have that $b^T\widehat{y} = \widehat{\gamma} := \min\{\|\widehat{W}x\|_1 : Ax = b\}$. Therefore, $(\widehat{s}, \widehat{y}, \widehat{w}, \widehat{u}, \widehat{v}, \widehat{\gamma})$ is a feasible point to the problem (2.11). Since $(s^*, y^*, w^*, u^*, v^*, \gamma^*)$ is an optimal solution to (2.11), by optimality, we have

$$\|\widehat{s}\|_0 \leq \|s^*\|_0. \tag{2.13}$$

Let x^* be an optimal solution to the problem $\min\{\|W^*x\|_1 : Ax = b\}$ where $W^* = \mathrm{diag}(w^*)$. Then by Lemma 2.6 again, $(t^*, x^*, \alpha^*, \beta^*)$, where $t^* = |x^*|, \alpha^* = |x^*| - x^*$ and $\beta^* = |x^*| + x^*$, is an optimal solution to the problem (2.8) with $w = w^*$. Note that $(s^*, y^*, w^*, u^*, v^*, \gamma^*)$ satisfies the constraints of the problem (2.11). Since $b^T y^* = \gamma^* = \min\{\|W^*x\|_1 : Ax = b\}$, by duality, (s^*, y^*, u^*, v^*) is an optimal solution to the dual problem (2.9) with $w = w^*$. By Lemma 2.6, the vectors s^* and $t^* = |x^*|$ are complementary, i.e., $(t^*)^T s^* = 0$. This implies that $\|t^*\|_0 + \|s^*\|_0 \leq n$. Combining this fact with (2.12) and (2.13) yields

$$\|x^*\|_0 = \|t^*\|_0 \leq n - \|s^*\|_0 \leq n - \|\widehat{s}\|_0 = \|\widehat{x}\|_0.$$

Thus x^* is a sparsest solution to the linear system (since $\widehat{x}$ is a sparsest solution). □

The above result implies that seeking the sparsest solutions of linear systems can be achieved by finding the densest slack variable $s \in R^n_+$ of the dual problem of the weighted ℓ_1-problem with all possible choices of $w \in R^n_+$. In fact, let $x(w)$ be an optimal solution to the weighted ℓ_1-problem (2.1), and let $(t(w), x(w), \alpha(w), \beta(w))$ be an optimal solution to the problem (2.8), and let $(s(w), y(w), u(w), v(w))$ be an optimal solution to the problem (2.9). By complementarity (Lemma 2.6), $s(w)$ and $t(w)$ are complementary. Thus $\|s(w)\|_0 + \|t(w)\|_0 \leq n$ for any given $w \in R^n_+$. By Lemma 2.6, we have $t(w) = |x(w)|$, and thus

$$\|s(w)\|_0 + \|x(w)\|_0 \leq n \text{ for any } w \in R^n_+. \tag{2.14}$$

By Corollary 2.5, for any sparsest solution x^*, there exists a weight w^* such that $x^* = x(w^*)$ is the unique solution to the weighted ℓ_1-problem (2.1) with $w = w^*$. By picking a strict complementarity solution $(s(w^*), y(w^*), u(w^*), v(w^*))$ to its dual problem (2.9), the equality can be achieved in (2.14). As a result, $s(w^*)$ is the densest vector among all possible choices of $w \in R^n_+$, i.e., $s(w^*) = \arg\max\{\|s(w)\|_0 : w \in R^n_+\}$. We summarize these facts as follows.

Corollary 2.8. *Let x^* be a sparsest solution of the system $Ax = b$. Then there exists a weight $w^* \in R^n_+$ such that the dual problem (2.9), where $w = w^*$, admits a solution (s^*, y^*, u^*, v^*) satisfying that $\|x^*\|_0 = n - \|s^*\|_0$.*

The weight w^* in Theorem 2.7 and Corollary 2.8 is referred to as an optimal weight in this paper. The above discussion indicates that an optimal weight can be found, in theory, by searching for the densest dual variable $s(w)$ among all possible choices of $w \in R^n_+$. It is the strict complementarity theory of linear programs that provides this new perspective to understand ℓ_0-minimization, leading to a new computational method for this problem.

3. A new reweighted ℓ_1-method. Theorem 2.7 indicates that solving the bilevel optimization problem (2.11) yields an optimal weight by which the sparsest solution of a system of linear equations can be found. However, directly solving a bilevel optimization problem to optimality is difficult. This motivates us to consider an approximation of (2.11). Let $\gamma(w)$ denote the optimal value of (2.1), i.e.,

$$\gamma(w) = \min_x\{\|Wx\|_1 : Ax = b\}, \tag{3.1}$$

where $W = \mathrm{diag}(w)$. By the duality theory of linear programs, the constraints of (2.11) imply that for any feasible point (y, s, w, u, v, γ) of (2.11), (y, s, u, v) must be an optimal solution to the problem (2.9). Thus the purpose of the bilevel problem (2.11) is actually to achieve two-level maximization. At the lower level, the dual objective $b^T y$ is maximized subject to the constraints of (2.9) for every given $w \geq 0$. This yields a feasible point, denoted by $(y(w), s(w), w, u(w), v(w), \gamma(w))$, to the bilevel problem (2.11). Then at the higher level, $\|s(w)\|_0$ is maximized among all possible choices of $w \geq 0$. Thus the following model is a certain approximation of (2.11):

$$\max\{\alpha\|s\|_0 + b^T y : \ A^T y - u + v = 0, \ s = w - u - v, \ (s, u, v, w) \geq 0\}, \tag{3.2}$$

where $\alpha > 0$ is a given parameter. The model (3.2) is used to maximize the combination of $\|s\|_0$ and $b^T y$ in order to possibly achieve the above-mentioned two-level maximization. By the structure of (2.11), the maximization of $\|s\|_0$ should be carried out under the constraint that (y, s, u, v) is an optimal solution to (2.9). This suggests that the parameter α in (3.2) should be chosen small.

Note that the function $\|s\|_0$ over the first orthant R^n_+ can be approximated by various concave functions (see [44, 42, 59]) which are called the merit functions for sparsity. There exists a class of merit functions for sparsity that are continuously differentiable in an open set containing R^n_+. Let $\Phi_\varepsilon : \mathcal{D} \to R_+$, where $\mathcal{D}$ is an open set containing R^n_+, be such a concave function satisfying that for every given $s \in R^n_+$, $\Phi_\varepsilon(s) \to \|s\|_0$ as $\varepsilon \to 0$. It is easy to construct such functions (see Definition 3.2 and Proposition 3.3 in this section and Examples 2.3–2.6 in [59]). Replacing $\|s\|_0$ by $\Phi_\varepsilon(s)$ in (3.2) yields

$$\max\{\alpha\Phi_\varepsilon(s) + b^T y : \ A^T y - u + v = 0, \ s = w - u - v, \ (s, u, v, w) \geq 0\}, \tag{3.3}$$

which is a convex optimization problem.

Note that the solution to the weighted ℓ_1-problem (2.1) is invariant when w is replaced by λw for any positive number $\lambda > 0$. We also note that if (y, s, w, u, v, γ) is an optimal solution to (2.11), then $\lambda(y, s, w, u, v, \gamma)$ is also an optimal solution to (2.11) for any positive number $\lambda > 0$, due to the fact $\|\lambda s\|_0 = \|s\|_0$. Therefore, on one hand, any weight that is large in magnitude can be scaled down to a small weight without affecting the optimal solution of (2.1) and the optimal objective value of (2.11). Thus w can be confined to a bounded convex set $\Omega \subset R^n_+$. On the other hand, the value of γ is not essential in (2.11) since (by a suitable scaling) the original strong-duality-type constraint $b^T y = \gamma = \min\{\|Wx\|_1 : Ax = b\}$ can be replaced by

$$b^T y = 1 = \min\{\|Wx\|_1 : Ax = b\} \tag{3.4}$$

without any damage of the conclusion in Theorem 2.7. However, this key constraint in model (2.11) is lost in (3.2) and (3.3). To achieve a good approximation of (2.11), we should include at least a certain relaxation of (3.4) into the model (3.3). Clearly, the weak duality condition is a judicious choice. Let $\gamma(w)$ be scaled down to 1 (under a suitable scaling of w). By the weak duality of linear programs, this is equivalent to imposing the same upper bound on the dual objective $b^T y$. Thus the constraint $b^T y \leq 1$ (as a relaxation of the strong-duality condition (3.4)), together with the constraint $w \in \Omega$, can be introduced into (3.3), leading to the well-defined model

$$\begin{aligned} &\max \alpha\Phi_\varepsilon(s) + b^T y \\ &\text{s.t. } A^T y - u + v = 0, \ s = w - u - v, \ b^T y \leq 1, \ w \in \Omega, \ (s, u, v, w) \geq 0. \end{aligned} \tag{3.5}$$

which admits a finite optimal value.

We now describe an algorithm based on (3.5). For simplicity, we fix a small $\varepsilon \in (0,1)$ and let α decrease iteratively in the course of the algorithm.

Algorithm 3.1. Let $\alpha_0 \in (0,1)$ be a given constant. Let $0 < \alpha^* \ll \alpha_0$ be a prescribed tolerance. Choose a bounded closed convex set $\Omega^0 \subseteq R^n_+$.

Step 1. If $\alpha_k \le \alpha^*$, stop; Otherwise, solve the convex optimization

$$\begin{aligned} &\max \alpha_k \Phi_\varepsilon(s) + b^T y \\ &\text{s.t. } A^T y - u + v = 0, s = w - u - v, b^T y \le 1, w \in \Omega^k, (w,s,u,v) \ge 0. \end{aligned} \tag{3.6}$$

Let $(w^{k+1}, y^{k+1}, s^{k+1}, u^{k+1}, v^{k+1})$ be a solution to this problem.

Step 2. Let $W^{k+1} = \mathrm{diag}(w^{k+1})$, and solve the weighted ℓ_1-problem

$$\gamma^{k+1} = \min\{\|W^{k+1}x\|_1 : \ Ax = b\} \tag{3.7}$$

to obtain a solution x^{k+1}.

Step 3. Choose $\alpha_{k+1} < \alpha_k$ and update Ω^k to obtain Ω^{k+1}. Then replace k by $k+1$ and return to Step 1.

The above algorithm may have a number of variants in terms of the updating schemes for α_k and Ω^k (ε also can be reduced iteratively). From a computational point of view, Ω^0 and Ω^k should be chosen as simple as possible. For instance, we may choose $\Omega^0 = \{w \in R^n_+ : \ \|w\|_1 \le \vartheta\}$, where ϑ is a given constant. More generally, we may pick any initial point $x^0 \in R^n$ and set

$$\Omega^0 = \{w \in R^n_+ : \ |x^0|^T w \le \vartheta, \ w \le \Gamma e\},$$

where $\vartheta > 0$ and $\Gamma > 0$ are given constants. We may fix $\Omega^k \equiv \Omega^0$ for all iterations, and may also change Ω^k iteratively in the course of algorithm. For example, based on the iterates (x^k, w^k, γ^k), Ω^k can be updated as

$$\Omega^k = \left\{w \in R^n_+ : \ |x^k|^T w \le \vartheta, \ w \le \Gamma^k e\right\}, \ \Gamma^k \ge \vartheta \|w^k\|_\infty / \gamma^k. \tag{3.8}$$

Such an update will be discussed later in sections 4 and 5. Also, we have a large number of choices for concave merit functions. For the convenience of our theoretical analysis, we are interested in the following class of functions.

Definition 3.2 ($\mathcal{M}$-class merit functions). *Let $\mathcal{M} := \{\Phi_\varepsilon\}$ be the set of merit functions for sparsity satisfying the following conditions: (i) for any given $s \in R^n_+$, $\Phi_\varepsilon(s) \to \|s\|_0$ as $\varepsilon \to 0$; (ii) $\Phi_\varepsilon(s)$ is continuously differentiable and concave with respect to s over an open set containing R^n_+; (iii) for any given constants $0 < c_1 < c_2$, there exists a small $\varepsilon^* > 0$ such that for any given $\varepsilon \in (0, \varepsilon^*]$,*

$$\Phi_\varepsilon(s) - \Phi_\varepsilon(\widehat{s}) \ge 1/2 \tag{3.9}$$

holds for any $0 \le s, \widehat{s} \le c_2 e$ satisfying that $\|\widehat{s}\|_0 < \|s\|_0$ and $c_1 \le s_i \le c_2$ for all $i \in supp(s)$.

The number $1/2$ in (3.9) is not essential and can be replaced by any fixed number in $(0,1)$. It is easy to construct a merit function in $\mathcal{M}$, as shown by the next proposition.

Proposition 3.3. *Let $\varepsilon \in (0,1)$. All the following functions are in the class $\mathcal{M}$:*

$$\Phi_\varepsilon(s) = \sum_{i=1}^{n} \left(1 - e^{-\frac{s_i}{\varepsilon}}\right), \quad \text{where } s \in R^n; \tag{3.10}$$

$$\Phi_\varepsilon(s) = \sum_{i=1}^{n} \frac{s_i}{s_i + \varepsilon}, \quad \text{where } s_i > -\varepsilon \text{ for all } i = 1, ..., n; \tag{3.11}$$

$$\Phi_\varepsilon(s) = n - \frac{1}{\log \varepsilon} \left(\sum_{i=1}^{n} \log(s_i + \varepsilon)\right), \text{ where } s_i > -\varepsilon \text{ for all } i = 1, ..., n. \tag{3.12}$$

Proof. It is straightforward to verify that all functions (3.10)–(3.12) satisfy conditions (i) and (ii) of Definition 3.2. It is also not very difficult to verify that these functions satisfy the condition (iii) of Definition 3.2. Let $s \in R^n_+$ be any vector satisfying $c_1 \le s_i \le c_2$ for all $i \in supp(s)$, where $0 < c_1 < c_2$ are two given constants, and let $\widehat{s}$ be any vector satisfying $0 \le \widehat{s} \le c_2 e$ and $\|\widehat{s}\|_0 < \|s\|_0$. Consider the function (3.10). We see that

$$\Phi_\varepsilon(\widehat{s}) = \sum_{\widehat{s}_j \neq 0} (1 - e^{-\frac{\widehat{s}_j}{\varepsilon}}) \le \|\widehat{s}\|_0.$$

Since $e^{-c_1/\varepsilon} \to 0$ as $\varepsilon \to 0$, there exists an $\varepsilon^* \in (0,1)$ such that $ne^{-\frac{c_1}{\varepsilon}} < 1/2$ for all $\varepsilon \in (0, \varepsilon^*]$. This implies that

$$\Phi_\varepsilon(s) = \sum_{s_j \neq 0} (1 - e^{-\frac{s_j}{\varepsilon}}) \ge \sum_{s_j \neq 0} (1 - e^{-\frac{c_1}{\varepsilon}}) = \|s\|_0 (1 - e^{-\frac{c_1}{\varepsilon}}) \ge \|s\|_0 - ne^{-\frac{c_1}{\varepsilon}} \ge \|s\|_0 - 1/2$$

for all $\varepsilon \in (0, \varepsilon^*]$. Therefore, $\Phi_\varepsilon(s) - \Phi_\varepsilon(\widehat{s}) \ge (\|s\|_0 - 1/2) - \|\widehat{s}\|_0 \ge 1/2$, where the last inequality follows from the fact $\|s\|_0 > \|\widehat{s}\|_0$ which implies that $\|s\|_0 - \|\widehat{s}\|_0 \ge 1$. So (3.10) satisfies condition (iii) of Definition 3.2. By a similar proof (the proof is omitted), we can verify that (3.11) and (3.12) satisfy the condition (iii) as well. □

By using a merit function in $\mathcal{M}$, the convex problem (3.6) can be solved efficiently by using existing gradient-type and interior-point-type methods. In the remainder of this paper, we address the following two questions: Under what condition do the iterates generated by Algorithm 3.1 converge to the sparsest solution of a system of linear equations? How good is the numerical performance of this algorithm, compared with some state-of-art sparsity-seeking methods?

4. Convergence analysis. For simplicity, we show the efficiency of the algorithm which adopts the updating scheme

$$\Omega^k \equiv \Omega, \ \alpha_{k+1} = \tau \alpha_k, \ k \ge 0, \tag{4.1}$$

where $\tau \in (0,1)$ is a given constant, and Ω is of the form

$$\Omega = \{w \in R^n_+ : \ |x^0|^T w \le \vartheta, \ w \le \Gamma e\}, \tag{4.2}$$

where $(\vartheta, \Gamma) > 0$ are given numbers, and x^0 is a given solution to the linear system.

At present, the guaranteed performance of various sparsity-seeking algorithms (such as the ℓ_1-method, reweighted ℓ_1-methods, greedy pursuits, and thresholding-type methods) has been analyzed mainly under the RIP, NSP or mutual-coherence-type assumptions which often imply the uniqueness of solutions to the ℓ_0-problems.

Our analysis is remarkably different from existing ones, and the results established in this section allow the ℓ_0-problem to possess multiple optimal solutions. We show that Algorithm 3.1 converges to a sparsest solution of the system of linear equations under some assumptions.

Let $\gamma(w)$ be defined by (3.1) and $\gamma^{\max}(\Omega)$ be the supremum of $\gamma(w)$ over Ω, i.e.,

$$\gamma^{\max}(\Omega) = \sup_{w\in\Omega} \gamma(w),$$

which is bounded above, as shown by the lemma below.

Lemma 4.1. *Consider the system $Ax = b$ ($\neq 0$). Let Ω be given by (4.2), where x^0 is a solution to the system $Ax = b$. Then*

$$0 < \gamma^{\max}(\Omega) \leq \vartheta. \tag{4.3}$$

In particular, let $w^0 \in R^n_{++}$ be any given vector and x^0 be a solution to the weighted ℓ_1-problem (2.1) with $w = w^0$, and let $\Gamma \geq \vartheta\|w^0\|_\infty/\gamma^0$, where $\gamma^0 = \gamma(w^0)$. Then $\gamma^{\max}(\Omega) = \vartheta$.

Proof. Note that $Ax^0 = b$. By the definition of Ω and the optimality, we see that $\gamma(w) \leq \|Wx^0\|_1 = |x^0|^T w \leq \vartheta$ for all $w \in \Omega$. Thus

$$\gamma^{\max}(\Omega) = \sup_{w\in\Omega} \gamma(w) \leq \vartheta. \tag{4.4}$$

Since $b \neq 0$, any solution to the linear system is nonzero, and hence $\gamma(w) > 0$ for any $w \in R^n_{++}$ in Ω. This implies that $\gamma^{\max}(\Omega) > 0$, which, together with (4.4), yields (4.3). In particular, for a given $w^0 \in R^n_{++}$, let x^0 and $\gamma^0 = \gamma(w^0)$ be an optimal solution and the optimal value to the weighted ℓ_1-problem (2.1) with $w = w^0$, respectively, and let $\Gamma \geq \vartheta\|w^0\|_\infty/\gamma^0$, which implies that $\vartheta w^0/\gamma^0 \leq \Gamma e$. By such choices of x^0 and Γ, we see that $\vartheta w^0/\gamma^0 \in \Omega$. Therefore, by the definition of $\gamma^{\max}(\Omega)$, we have

$$\vartheta = \gamma(\vartheta w^0/\gamma^0) \leq \gamma^{\max}(\Omega), \tag{4.5}$$

where the equality follows from the fact that $\gamma(\vartheta w^0/\gamma^0) = \vartheta\gamma(w^0/\gamma^0)$ and $\gamma(w^0/\gamma^0) = 1$. Combining (4.4) and (4.5) yields $\gamma^{\max}(\Omega) = \vartheta$, as desired. □

Throughout the remainder of this paper, we use $\mathcal{S}^*$ to denote the set of the sparsest solutions of the linear system $Ax = b$. Note that $A_{supp(x^*)}$ has a full-column rank for every $x^* \in \mathcal{S}^*$. Thus $\mathcal{S}^*$ contains only a finite number of elements. By Corollary 2.5, for every $x^* \in \mathcal{S}^*$, there exists a weight $w^* \in R^n_+$ such that x^* is the unique solution to the weighted ℓ_1-problem

$$\min\{\|W^*x\|_1 : Ax = b\} \tag{4.6}$$

to which the dual problem is given as

$$\max_{(y,s,u,v)} \left\{b^T y : \ A^T y - u + v = 0, \ s = w^* - u - v, \ (s,u,v) \geq 0\right\}. \tag{4.7}$$

As we have seen from section 2, there exist infinitely many optimal weights for every $x^* \in \mathcal{S}^*$. We denote by $\mathcal{Y}(x^*)$ the set of optimal weights for x^*, i.e.,

$$\mathcal{Y}(x^*) := \{w^* \in R^n_+ : \ x^* \text{ is the unique solution to the problem (4.6)}\}.$$

Let

$$\Omega^* = \bigcup_{x^*\in\mathcal{S}^*} \mathcal{Y}(x^*)$$

be the set of all optimal weights associated with the sparsest solutions of the linear system $Ax = b$. Note that the scaling of a weight does not change the solution of weighted ℓ_1-minimization. For every $w^* \in \mathcal{Y}(x^*)$, we have that $\lambda w^* \in \mathcal{Y}(x^*)$ for any $\lambda > 0$, and thus $\mathcal{Y}(x^*)$ and Ω^* are cones. Since $\lambda w^* \in \Omega$ for all sufficiently small $\lambda > 0$, we have that $\mathcal{Y}(x^*) \cap \Omega \neq \emptyset$ for every $x^* \in \mathcal{S}^*$, and hence $\Omega^* \cap \Omega \neq \emptyset$. Given $x^* \in \mathcal{S}^*$ and $w^* \in \mathcal{Y}(x^*)$, we define the set

$$\Upsilon(w^*, x^*) = \{s : \ (y, s, u, v) \text{ is an optimal solution to (4.7)}, \ |x^*|^T s = 0, \ |x^*| + s > 0\}.$$

Clearly, $s \in \Upsilon(w^*, x^*)$ if and only if there exist some vectors y, u, and v such that $(x^*, (y, s, u, v))$ is a pair of strictly complementary solutions to (4.6) and (4.7). Such a pair always exists by the linear programming theory. Thus, $\Upsilon(w^*, x^*) \neq \emptyset$ for any given $x^* \in \mathcal{S}^*$ and $w^* \in \mathcal{Y}(x^*)$.

Before showing our main convergence theorem, we state several technical results.

Lemma 4.2. *Let $\Phi_\varepsilon(s) \in \mathcal{M}$ (as specified in Definition 3.2). Let $x^* \in \mathcal{S}^*$ and $w^* \in \mathcal{Y}(x^*) \cap \Omega$. Then for any given $s^* \in \Upsilon(w^*, x^*)$, there exists a sufficiently small $\varepsilon^* \in (0, 1)$ accordingly such that for any $\varepsilon \in (0, \varepsilon^*]$, the inequality*

$$\Phi_\varepsilon(s^*) - \Phi_\varepsilon(s) \geq 1/2 \tag{4.8}$$

holds for any s satisfying $\|s\|_0 < \|s^*\|_0$ and $s \in T(\Omega, A)$ where

$$T(\Omega, A) := \left\{s : \ A^T y - u + v = 0, \ s = w - u - v, \ (s, u, v) \geq 0, \ w \in \Omega\right\}. \tag{4.9}$$

Proof. Note that the vector $s = w - u - v \leq w$ for any $u, v \geq 0$. Since Ω is bounded, the set $T(\Omega, A)$ given by (4.9) is also bounded. There exists an upper bound $c_2 > 0$ such that $0 \leq s \leq c_2 e$ for all $s \in T(\Omega, A)$. Let $w^* \in \mathcal{Y}(x^*) \cap \Omega$. By the definition of $\mathcal{Y}(x^*)$, x^* is the unique solution to the problem (4.6). Let s^* be a given vector in $\Upsilon(w^*, x^*)$. Then there exists a vector (y^*, u^*, v^*) such that (y^*, s^*, u^*, v^*) is a solution to the problem (4.7) and that $|x^*|$ and s^* are strictly complementary, i.e., $|x^*|^T s^* = 0$ and $|x^*| + s^* > 0$. Since x^* is a sparsest solution to the linear system, the matrix $A_{supp(x^*)} \in R^{m \times |supp(x^*)|}$ has a full-column rank. Thus $\|x^*\|_0 = |supp(x^*)| \leq m$. Since $m < n$ and $|x^*|$ and s^* are strictly complementary, the vector s^* must contain at least $n - m$ positive components. Thus we define

$$c_1 = \min_{s_i^* > 0} s_i^*,$$

which is a positive number. Since $w^* \in \mathcal{Y}(x^*) \cap \Omega$ and (y^*, s^*, u^*, v^*) is a solution to (4.7), it is easy to see that $s^* \in T(\Omega, A)$. Thus s^* satisfies that $0 \leq s^* \leq c_2 e$. By the definition of c_1, we see that $0 < c_1 \leq s_i^* \leq c_2$ for all $i \in supp(s^*)$. Since $\Phi_\varepsilon(s) \in \mathcal{M}$, for the above-defined c_1 and c_2, there exists a number $\varepsilon^* \in (0, 1)$ such that $\Phi_\varepsilon(s^*) - \Phi_\varepsilon(s) \geq 1/2$ holds for any $\varepsilon \in (0, \varepsilon^*]$ and for any s satisfying that $\|s\|_0 < \|s^*\|_0$ and $s \in T(\Omega, A)$. □

The next result claims that under some condition, an optimal solution of the problem (2.9) with a scaled weight can be constructed from a feasible solution of the original problem (2.9).

Lemma 4.3. *Let $w \in R^n_+$ be given. Consider the problem (2.1) and its dual problem (2.9). Suppose that $\gamma(w) > 1$. If $(\overline{y}, \overline{s}, \overline{u}, \overline{v})$ is a feasible solution to the problem (2.9) satisfying that $b^T \overline{y} = 1$ and $|\overline{u} - \overline{v}| \leq (\overline{u} + \overline{v})/\gamma(w)$, then $(y, s, u, v) = (\overline{y}, \frac{\overline{s}}{\gamma(w)}, u', v')$, where $u' = \frac{1}{2}[\overline{u} - \overline{v} + (\overline{u} + \overline{v})/\gamma(w)]$ and $v' = \frac{1}{2}[\overline{v} - \overline{u} + (\overline{u} + \overline{v})/\gamma(w)]$, is an optimal solution to the problem*

$$\max_{(y,s,u,v)} \{b^T y : \ A^T y - u + v = 0, \ s = \frac{w}{\gamma(w)} - u - v, \ (s, u, v) \geq 0\}. \tag{4.10}$$

Proof. Since $(\overline{y}, \overline{s}, \overline{u}, \overline{v})$ is a feasible solution to (2.9), we have that $A^T\overline{y} = \overline{u} - \overline{v}$, $w - \overline{s} = \overline{u} + \overline{v}$, and $(\overline{s}, \overline{u}, \overline{v}) \geq 0$. Since $|\overline{u} - \overline{v}| \leq (\overline{u} + \overline{v})/\gamma(w)$, we see that

$$u' = \frac{1}{2}\left[\overline{u} - \overline{v} + (\overline{u} + \overline{v})/\gamma(w)\right] \geq 0,\ v' = \frac{1}{2}\left[\overline{v} - \overline{u} + (\overline{u} + \overline{v})/\gamma(w)\right] \geq 0$$

and that

$$u' + v' = (\overline{u} + \overline{v})/\gamma(w) = (w - \overline{s})/\gamma(w),\ u' - v' = \overline{u} - \overline{v} = A^T\overline{y}.$$

Thus $(y, s, u, v) = (\overline{y}, \overline{s}/\gamma(w), u', v')$ is a feasible solution to the problem (4.10). Note that the problem (4.10) is the dual problem of the weighted ℓ_1-problem $\min\{\|(\frac{W}{\gamma(w)})x\|_1 : Ax = b\}$ to which the optimal value is 1. By strong duality, the optimal value of the dual problem (4.10) is also 1. Since $b^T\overline{y} = 1$, the feasible point $(y, s, u, v) = (\overline{y}, \overline{s}/\gamma(w), u', v')$ is an optimal solution to the problem (4.10). □

We will also make use of the following perturbation theorem of linear programs.

Lemma 4.4 (Mangasarian and Meyer [45]). *Consider the linear program* $\min\{c^T u : u \in Q\}$, *where* $Q \subseteq R^n$ *is the feasible set. Let* f *be a continuously differentiable convex function on some open set containing* Q. *If the solution set* $\bar{S}$ *of this linear program is nonempty and bounded, and* $c^T u + \widetilde{\alpha} f(u)$ *is bounded from below on* Q *for some* $\widetilde{\alpha} > 0$, *then the solution set of the perturbed problem*

$$\min\{c^T u + \alpha f(u) :\ u \in Q\}$$

is contained in $\bar{S}$ *for sufficiently small* $\alpha > 0$.

To prove the convergence of Algorithm 3.1, we need to impose some assumptions on the linear system $Ax = b$. Define the set

$$\widetilde{\Omega^*} := \{w^* \in \Omega^* :\ \gamma(w^*) = 1\}$$

which is a nonempty subset of Ω^*. The nonemptiness of $\widetilde{\Omega^*}$ follows from the fact that Ω^* is a cone. In fact, by scaling if necessary, there is a vector $w^* \in \Omega^*$ such that $\gamma(w^*) = 1$. We impose the following assumption on the problem.

Assumption 4.5. $\Omega \cap \widetilde{\Omega^*} \neq \emptyset$.

Later, we will show that Assumption 4.5 is satisfied when ϑ and Γ in (4.2) are suitably chosen (see Lemma 4.7 for details). The main convergence theorem is summarized as follows.

Theorem 4.6. *Consider Algorithm 3.1 with the updating scheme (4.1). Let* $\Phi_\varepsilon \in \mathcal{M}$ *and* Ω *be given as (4.2). Suppose that Assumption 4.5 is satisfied. Then there exist a sufficiently small parameter* $\varepsilon^* \in (0, 1)$ *and a sufficiently small tolerance* $\alpha^* \in (0, 1)$ *such that for any fixed small* $\varepsilon \in (0, \varepsilon^*]$, *the sequence* $\{(x^k, w^k, u^k, v^k)\}$, *generated by Algorithm 3.1, satisfies the following properties:*

(i) *For* $\vartheta = 1$, *after* $k_0 := \lceil \log(\alpha^*/\alpha_0)/\log\tau \rceil$ *steps, the iterate* x^k *(*$k \geq k_0$*) must be a sparsest solution to the system* $Ax = b$.

(ii) *For* $\vartheta > 1$, *after* $k_0 := \lceil \log(\alpha^*/\alpha_0)/\log\tau \rceil$ *steps, if* $|u^k - v^k| \leq (u^k + v^k)/\gamma(w^k)$ *holds for a* $k \geq k_0$, *then* x^k *must be a sparsest solution to the system* $Ax = b$.

Proof. By Assumption 4.5, there exists a vector $w^* \in \Omega \cap \widetilde{\Omega^*}$. This vector satisfies that $w^* \in \Omega^*,\ w^* \in \Omega$, and $\gamma(w^*) = 1$. By the definition of Ω^*, there exists a sparsest solution $x^* \in \mathcal{S}^*$ such that $w^* \in \mathcal{Y}(x^*)$. Also, there exists a vector $s^* \in \Upsilon(w^*, x^*)$. Given such vectors (w^*, x^*, s^*), by Lemma 4.2, there exists a sufficiently small $\varepsilon^* > 0$ accordingly so that for any given $\varepsilon \in (0, \varepsilon^*]$, the inequality

$$\Phi_\varepsilon(s^*) - \Phi_\varepsilon(s) \geq 1/2 \tag{4.11}$$

holds for all s satisfying $\|s\|_0 < \|s^*\|_0$ and $s \in T(\Omega, A)$ defined by (4.9)

We now consider the problem (3.6) with such a fixed $\varepsilon \in (0, \varepsilon^*]$. For a given $\alpha_k \in (0,1)$, the problem (3.6) is a convex optimization problem, which can be viewed as a perturbed version of the linear program

$$\begin{aligned} \max_{(w,y,s,u,v)} \quad & b^T y \\ \text{s.t.} \quad & A^T y - u + v = 0, s = w - u - v, b^T y \le 1, (w,s,u,v) \ge 0, w \in \Omega. \end{aligned} \tag{4.12}$$

We use $\mathcal{D}^*$ to denote the set of optimal solutions of (4.12). Note that $A \in R^{m\times n}(m < n)$ has a full-row rank. Since Ω is bounded, the feasible set of (4.12) is bounded, and so is the solution set $\mathcal{D}^*$. In fact, the constraints of (4.12) imply that

$$0 \le s \le w,\ 0 \le u \le w,\ 0 \le v \le w,\ y = (AA^T)^{-1}A(u-v),$$

and hence

$$\|y\|_\infty \le \|(AA^T)^{-1}A\|_\infty \|u-v\|_\infty \le 2\|(AA^T)^{-1}A\|_\infty \|w\|_\infty.$$

Thus the boundedness of Ω implies that the feasible set (and hence the solution set $\mathcal{D}^*$) of (4.12) is bounded. By Definition 3.2, $\Phi_\varepsilon(s)$ is a continuously differentiable concave function over an open neighborhood of the first orthant $\{s \in R^n : s \ge 0\}$. Thus the function

$$\overline{\Phi}_\varepsilon(w,y,s,u,v) := \Phi_\varepsilon(s) + 0^T w + 0^T y + 0^T u + 0^T v$$

is a continuously differentiable concave function over an open neighborhood of the feasible set of (4.12). Since the feasible set is bounded, for any given $\alpha > 0$, the concave function

$$\alpha\Phi_\varepsilon(s) + b^T y = \alpha\widetilde{\Phi}_\varepsilon(w,y,s,u,v) + b^T y$$

is bounded from above over the feasible set of (4.12). By Lemma 4.4, there exists a sufficiently small number $\alpha^* > 0$ such that for any $\alpha_k \in (0, \alpha^*]$, the solution set of (3.6) is contained in $\mathcal{D}^*$. Thus the optimal solution $(w^{k+1}, y^{k+1}, s^{k+1}, u^{k+1}, v^{k+1})$ to the problem (3.6) is also an optimal solution to the problem (4.12) when $\alpha_k \in (0, \alpha^*]$. By the updating scheme (4.1), α_k is reduced by a factor $\tau < 1$ at each iteration. So after a finite number of iterations, i.e., $k \ge k_0 =: \lceil \log(\alpha^*/\alpha_0)/\log\tau \rceil$, we must have that $\alpha_k \in (0, \alpha^*]$.

We now prove that $b^T y^{k+1} = 1$ for $k \ge k_0$. On one hand, the constraint $b^T y \le 1$ in (4.12) implies that

$$b^T y^{k+1} \le 1. \tag{4.13}$$

On the other hand, for the vector (w^*, x^*, s^*) specified at the beginning of this proof, there exists vectors (y^*, u^*, v^*) accordingly such that (y^*, s^*, u^*, v^*) is an optimal solution to the dual problem (4.7), and that $|x^*|$ and s^* are strictly complementary. Since $w^* \in \widehat{\Omega^*}$, by strong duality, we have that $b^T y^* = \gamma(w^*) = 1$. Therefore $(w^*, y^*, s^*, u^*, v^*)$ is a feasible solution to the problem (4.12). By optimality, we must have that

$$b^T y^{k+1} \ge b^T y^* = 1. \tag{4.14}$$

Merging (4.13) and (4.14) yields

$$b^T y^{k+1} = 1 \text{ for all } k \geq k_0. \tag{4.15}$$

We now consider the weighted ℓ_1-problem with $W^{k+1} = \text{diag}(w^{k+1})$

$$\gamma^{k+1} = \min_x \{\|W^{k+1}x\|_1 : \ Ax = b\} \tag{4.16}$$

and its dual problem

$$\max_{(y,s,u,v)} \{b^T y : A^T y - u + v = 0, \ s = w^{k+1} - u - v, \ (s,u,v) \geq 0\}. \tag{4.17}$$

By the construction of Step 2 of Algorithm 3.1, we see that $w^{k+1} \in \Omega$. Note that $(y^{k+1}, s^{k+1}, u^{k+1}, v^{k+1})$ is a feasible point to the problem (4.17) and it satisfies (4.15). The optimal value of (4.17) is at least 1. By strong duality, the optimal value of the problem (4.16) is also at least 1, i.e.,

$$\gamma^{k+1} \geq 1. \tag{4.18}$$

We now consider the following two cases.

Case 1: $\vartheta = 1$. In this case, by Lemma 4.1, we have $\gamma^{k+1} = \gamma(w^{k+1}) \leq \gamma^{\max}(\Omega) \leq \vartheta = 1$. This, together with (4.18), implies that $\gamma(w^{k+1}) = 1$. By strong duality again, this in turn implies that the optimal value of the dual problem (4.17) is also 1. Since $b^T y^{k+1} = 1$, we deduce that $(y^{k+1}, s^{k+1}, u^{k+1}, v^{k+1})$ is an optimal solution to the problem (4.17). Thus, by Lemma 2.6, s^{k+1} and $|x^{k+1}|$ are complementary, i.e.,

$$|x^{k+1}|^T s^{k+1} = 0 \tag{4.19}$$

where x^{k+1} is a solution to the problem (4.16).

Case 2: $\vartheta > 1$ and $|u^{k+1} - v^{k+1}| \leq (u^{k+1} + v^{k+1})/\gamma^{k+1}$ for some $k \geq k_0$. If (4.18) holds as an equality, the same proof in Case 1 yields (4.19). Thus it is sufficient to consider the case $\gamma^{k+1} > 1$. Note that $(y^{k+1}, s^{k+1}, u^{k+1}, v^{k+1})$ is a feasible solution to (4.17) with $b^T y^{k+1} = 1$ for $k \geq k_0$. Let

$$u' = \frac{1}{2}[u^{k+1} - v^{k+1} + (u^{k+1} + v^{k+1})/\gamma^{k+1}], v' = \frac{1}{2}[v^{k+1} - u^{k+1} + (u^{k+1} + v^{k+1})/\gamma^{k+1}].$$

Then by Lemma 4.3, $(y, s, u, v) := (y^{k+1}, s^{k+1}/\gamma^{k+1}, u', v')$ is an optimal solution to the problem

$$\max_{(y,s,u,v)} \{b^T y : A^T y - u + v = 0, \ s = \frac{w^{k+1}}{\gamma^{k+1}} - u - v, \ (s,u,v) \geq 0\},$$

which is the dual problem of the weighted ℓ_1-problem

$$\min_x \left\{ \left\| \left(\frac{W^{k+1}}{\gamma^{k+1}} \right) x \right\|_1 : \ Ax = b \right\}, \tag{4.20}$$

where $W^{k+1} = \text{diag}(w^{k+1})$. Since the scaling of weight does not affect the solution to the problem (4.16), x^{k+1} remains an optimal solution to the problem (4.20). By Lemma 2.6 again, we deduce that $|x^{k+1}|$ and $\frac{s^{k+1}}{\gamma^{k+1}}$ are complementary, and thus (4.19) remains valid.

Therefore, for both cases 1 and 2 above, we have that

$$\|x^{k+1}\|_0 + \|s^{k+1}\|_0 \le n. \tag{4.21}$$

Note that $|x^*|$ and s^* are strictly complementary. This implies that

$$\|x^*\|_0 + \|s^*\|_0 = n. \tag{4.22}$$

We now prove that for $k \ge k_0$, x^{k+1} is a sparsest solution to system $Ax = b$. Assume the contrary—that x^{k+1} is not a sparsest solution, i.e., $\|x^{k+1}\|_0 > \|x^*\|_0$. Then combining (4.21) and (4.22) yields

$$\|s^*\|_0 - \|s^{k+1}\|_0 \ge (n - \|x^*\|_0) - (n - \|x^{k+1}\|_0) = \|x^{k+1}\|_0 - \|x^*\|_0 > 0.$$

Note that $s^* \in \Upsilon(w^*, x^*)$ and $s^{k+1} \in T(\Omega, A)$ for $k \ge k_0$ (since $(w^{k+1}, y^{k+1}, s^{k+1}, u^{k+1}, v^{k+1})$ is an optimal solution to (4.12) for $k \ge k_0$). It follows from Lemma 4.2 that

$$\Phi_\varepsilon(s^*) - \Phi_\varepsilon(s^{k+1}) \ge 1/2 \tag{4.23}$$

holds for the fixed small $\varepsilon \in (0, \varepsilon^*]$ and for any s^{k+1} with $k \ge k_0$. On the other hand, since $(w^*, y^*, s^*, u^*, v^*)$ is a feasible point to (3.6), and since $(w^{k+1}, y^{k+1}, s^{k+1}, u^{k+1}, v^{k+1})$ is an optimal solution to (3.6) for $k \ge k_0$, by optimality, we must have that

$$\alpha_k \Phi_\varepsilon(s^{k+1}) + b^T y^{k+1} \ge \alpha_k \Phi_\varepsilon(s^*) + b^T y^*, \ k \ge k_0.$$

By (4.15) and $b^T y^* = 1$, the above inequality is reduced to $\Phi_\varepsilon(s^{k+1}) - \Phi_\varepsilon(s^*) \ge 0$ for all sufficiently large $k \ge k_0$. This contradicts (4.23). Therefore, for sufficiently large k, x^{k+1} generated at Step 2 of Algorithm 3.1 is a sparsest solution to the linear system $Ax = b$. □

We now prove that Assumption 4.5 is satisfied, roughly speaking, if ϑ and Γ are suitably chosen. Given the system $Ax = b$, let $\sigma^*(A, b)$ denote the constant

$$\sigma^*(A, b) = \min_{x^* \in \mathcal{S}^*} \left\| A^T_{J_0(x^*)} A_{supp(x^*)} (A^T_{supp(x^*)} A_{supp(x^*)})^{-1} \right\|_\infty, \tag{4.24}$$

where $\mathcal{S}^*$ is the set of the sparsest solutions of the linear system $Ax = b$.

Lemma 4.7. *Let $w^0 \in R^n_{++}$ be a given vector and x^0 be an optimal solution of the problem $\min\{\|W^0 x\|_1 : Ax = b\ (\neq 0)\}$ to which the optimal value is $\gamma^0 = \gamma(w^0)$. Choose (ϑ, Γ) satisfying that*

$$\vartheta \ge 1, \ \vartheta > \beta \sigma^*(A, b), \ \Gamma \ge \vartheta \|w^0\|_\infty / \gamma^0, \tag{4.25}$$

where $\sigma^*(A, b)$ is the constant defined by (4.24) and $\beta = \frac{\|w^0\|_\infty}{\min_{1 \le i \le n} w^0_i}$. Then the set $\Omega = \{w \in R^n_+ : \ |x^0|^T w \le \vartheta, \ w \le \Gamma e\}$ satisfies Assumption 4.5, i.e., $\Omega \cap \widetilde{\Omega^*} \neq \emptyset$.

Proof. Let $w^0 \in R^n_{++}$ be a given vector, and let ϑ and Γ be chosen to satisfy (4.25). Consider the weighted ℓ_1-problem

$$\min\{\|W^0 x\|_1 : \ Ax = b\}. \tag{4.26}$$

Since $b \neq 0$ and $w^0 \in R^n_{++}$, the optimal value $\gamma^0 = \gamma(w^0)$ is positive. Define $\widehat{w} = w^0/\gamma^0$. Note that the scaling of a weight does not change the solution of (4.26). So x^0 remains an optimal solution to the problem

$$\gamma(\widehat{w}) = \min\{\|\widehat{W} x\|_1 : \ Ax = b\}, \tag{4.27}$$

where $\widehat{W} = \mathrm{diag}(\widehat{w})$. Clearly, $\gamma(\widehat{w}) = 1$. By (4.24) and finiteness of $\mathcal{S}^*$, there exists a sparsest solution $x^* \in \mathcal{S}^*$ such that

$$\left\| A^T_{J_0(x^*)} A_{supp(x^*)} (A^T_{supp(x^*)} A_{supp(x^*)})^{-1} \right\|_\infty = \sigma^*(A,b) < \vartheta/\beta. \tag{4.28}$$

where the inequality follows from (4.25). In what follows, we use J to denote $supp(x^*)$ and J_c to denote $J_0(x^*)$ for simplicity. Let θ be defined as

$$\theta := \vartheta \|\widehat{W} x^*\|_1 = \vartheta (\widehat{w}_J)^T |x^*_J| \geq \vartheta \gamma(\widehat{w}) = \vartheta,$$

where the inequality follows from the fact that $\gamma(\widehat{w})$ is the minimum value of (4.27). We now construct a weight $w^* \in R^n_{++}$ so that x^* is the unique solution to the problem

$$\gamma(w^*) = \min\{\|W^* x\|_1 : \ Ax = b\}, \tag{4.29}$$

where $W^* = \mathrm{diag}(w^*)$. In fact, we may define w^* as follows:

$$w^*_J = (\frac{\vartheta}{\theta})\widehat{w}_J, \ w^*_{J_c} = \left(\frac{\vartheta(\sigma^*(A,b) + \delta)\|\widehat{w}_J\|_\infty}{\theta} \right) e_{J_c}, \tag{4.30}$$

where $\delta > 0$ is a positive number given by $\delta = \vartheta/\beta - \sigma^*(A,b)$. Note that $\||B|\|_\infty = \|B\|_\infty$ for any matrix B. By (4.28) and (4.30), we have

$$\left\| |A^T_{J_c} A_J (A^T_J A_J)^{-1} | w^*_J \right\|_\infty \leq \left\| |A^T_{J_c} A_J (A^T_J A_J)^{-1}| \right\|_\infty \|w^*_J\|_\infty < [\sigma^*(A,b) + \delta] \| \frac{\vartheta \widehat{w}_J}{\theta} \|_\infty.$$

The above inequalities, together with (4.30), imply that

$$\left| A^T_{J_c} A_J (A^T_J A_J)^{-1} \right| w^*_J < (\sigma^*(A,b) + \delta)(\vartheta \|\widehat{w}_J\|_\infty/\theta) e_{J_c} = w^*_{J_c}.$$

Therefore, by Corollary 2.5, x^* is the unique solution to the problem (4.29) where w^* is determined by (4.30), and thus $w^* \in \mathcal{Y}(x^*) \subseteq \Omega^*$. Moreover, the optimal value of (4.29) is given by

$$\gamma(w^*) = \|W^* x^*\|_1 = (w^*_J)^T |x^*_J| = \vartheta (\widehat{w}_J)^T |x^*_J|/\theta = 1.$$

Therefore, $w^* \in \widetilde{\Omega^*} = \{w \in \Omega^* : \gamma(w) = 1\}$. To show that $\Omega \cap \widetilde{\Omega^*} \neq \emptyset$, it suffices to show that w^* is also in Ω. Indeed, note that $\|\widehat{w}_J\|_\infty \leq \|w^0\|_\infty/\gamma^0 = \beta \left(\min_{1 \leq i \leq n} w^0_i\right)/\gamma^0$, which implies that

$$|x^0_{J_c}|^T (\|\widehat{w}_J\|_\infty e_{J_c}) \leq \beta |x^0_{J_c}|^T \left[\left(\min_{1 \leq i \leq n} w^0_i \right) e_{J_c} \right] / \gamma^0 \leq \beta |x^0_{J_c}|^T w^0_{J_c} / \gamma^0 = \beta |x^0_{J_c}|^T \widehat{w}_{J_c}.$$

This inequality, with $\beta(\sigma^*(A,b) + \delta) = \vartheta \geq 1, \theta \geq \vartheta$ and $|x^0|^T \widehat{w} = \gamma(\widehat{w}) = 1$, implies that

$$\begin{aligned}
|x^0|^T w^* &= \vartheta |x^0_J|^T \widehat{w}_J/\theta + \vartheta(\sigma^*(A,b) + \delta) |x^0_{J_c}|^T \left(\|\widehat{w}_J\|_\infty e_{J_c} \right)/\theta \\
&\leq \vartheta |x^0_J|^T \widehat{w}_J/\theta + \vartheta(\sigma^*(A,b) + \delta)\beta |x^0_{J_c}|^T \widehat{w}_{J_c}/\theta \\
&= \vartheta |x^0_J|^T \widehat{w}_J/\theta + \vartheta^2 |x^0_{J_c}|^T \widehat{w}_{J_c}/\theta \\
&\leq \vartheta^2 (|x^0|^T \widehat{w})/\theta = \vartheta^2/\theta \leq \vartheta
\end{aligned}$$

and

$$\begin{aligned}\|w^*\|_\infty &= \max\{\|w^*_J\|_\infty, \|w^*_{J_c}\|_\infty\} = \vartheta \max\{\|\widehat{w}_J\|_\infty, (\sigma^*(A,b)+\delta)\|\widehat{w}_J\|_\infty\}/\theta \\ &= (\vartheta/\theta)\max\{1, (\sigma^*(A,b)+\delta)\}\|\widehat{w}_J\|_\infty \\ &\le \max\{1, (\sigma^*(A,b)+\delta)\}(\|w^0\|_\infty/\gamma^0) \le \vartheta\|w^0\|_\infty/\gamma^0 \le \Gamma,\end{aligned}$$

where the first inequality follows from the fact $\vartheta/\theta \le 1$ and $\|\widehat{w}_J\|_\infty \le \|w^0\|_\infty/\gamma^0$, the second inequality follows from $\sigma^*(A,b)+\delta \le (\sigma^*(A,b)+\delta)\beta = \vartheta$, and the final inequality follows from (4.25). Therefore, $w^* \in \Omega$, as desired. □

Remark 4.1. It is worth stressing that Assumption 4.5 is a mild condition, as indicated by Lemma 4.7. The existing assumptions for sparsity-seeking algorithms such as the mutual coherence, RIP, NSP, and ERC are not explicitly needed in our convergence analysis. However, this analysis indicates that for a given linear system, the constant $\sigma^*(A,b)$ defined by (4.24) provides an important clue for the sparsest solutions of linear systems. Based on this constant, Assumption 4.5 will be satisfied when the parameters (ϑ, Γ) are chosen large enough. In particular, when $\sigma^*(A,b) < 1$, Lemma 4.7 indicates that the parameter $\vartheta = 1$ can be taken provided that w^0 is drawn in the neighborhood of e. In this case, Algorithm 3.1 with $\vartheta = 1$ guarantees finding the sparsest solution of linear systems, as shown by Theorem 4.6. The condition $\sigma^*(A,b) < 1$ can be guaranteed when the mutual coherence condition or the more general ERC condition is satisfied. In fact, it has been shown in [52] that the mutual coherence condition $\|x^*\|_0 < \frac{1}{2}(1 + \frac{1}{\mu(A)})$ implies that the ERC condition $\|A^\dagger_{supp(x^*)}A_{J_0(x^*)}\|_1 < 1$, where $A^\dagger_{supp(x^*)} = (A^T_{supp(x^*)}A_{supp(x^*)})^{-1}A^T_{supp(x^*)}$. Clearly, in terms of $\sigma^*(A,b)$, the ERC condition is equivalent to $\sigma^*(A,b) < 1$. As a result, it follows from Lemma 4.7 and Theorem 4.6 that Algorithm 3.1 converges to the sparsest solution of linear systems if the mutual coherence condition or the ERC is satisfied. The mutual coherence condition implies that x^* is the unique sparsest solution to the linear system, but the ERC does not, as shown by the example

$$A = \begin{bmatrix} \frac{1}{2\sqrt{2}} & 0 & -1 & -\frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2\sqrt{2}} & -1 & -\frac{1}{2} & -\frac{1}{\sqrt{2}} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & -1 & \frac{1}{2} & 0 & 0 & 0 \end{bmatrix}, \; b = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}. \tag{4.31}$$

Clearly, the system $Ax = b$ admits three sparsest solutions: $x^1 = (0,0,0,0,1,0,0)^T$, $x^2 = (0,0,0,0,0,\sqrt{2},0)^T$, and $x^3 = (0,0,0,0,0,0,-\sqrt{2})^T$. However, for this example,

$$\begin{aligned}\sigma^*(A,b) &= \left\| A^T_{J_0(x^1)}A_{supp(x^1)}(A^T_{supp(x^1)}A_{supp(x^1)})^{-1} \right\|_\infty \\ &= \left\| \left(\frac{1}{4}, \frac{1}{4}, 0, 0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \right\|_\infty = \frac{1}{\sqrt{2}} < 1.\end{aligned}$$

Thus this example satisfies the ERC. Given $w^0 > 0$ with $\beta = \|w^0\|_\infty / \min_{1\le i\le n} w^0_i$, we may choose $\vartheta \ge 1, \vartheta > \beta/\sqrt{2}$ and $\Gamma = \vartheta\|w^0\|_\infty/\gamma^0$ to satisfy (4.25), which ensures that Assumption 4.5 is satisfied. Moreover, $\vartheta = 1$ can be taken provided w^0 is chosen so that $\beta < \sqrt{2}$. This example also shows that the convergence of Algorithm 3.1 does not require the uniqueness of the sparsest solutions of linear systems. This is remarkably different from existing efficiency analyses for sparsity-seeking methods, which are often carried out under such conditions as the mutual coherence, RIP, or NSP of order $2k$. These conditions imply the uniqueness of k-sparse solutions of linear systems. It is easy to see that the example (4.31) does not satisfy the RIP or NSP of

order 2. Thus Assumption 4.5 is mild in the sense that it can be met by choosing the parameters (ϑ, Γ) large enough, despite the situations where the linear system may admit multiple sparsest solutions, to which the existing RIP- or NSP-based analysis doest not apply.

Remark 4.2. Clearly, under the scheme (4.1), performing r iterations of Algorithm 3.1 is equivalent to setting $\alpha = (\tau)^r \alpha_0$ and using this α to perform only one iteration of Algorithm 3.1. Thus a simple implementation of Algorithm 3.1 is to fix $(\alpha, \varepsilon, \Omega)$ and perform only one iteration, leading to the so-called NewRW-Heu method in section 5. The numerical performance of this method is demonstrated in section 5, and the theoretical efficiency of this method can follow directly from Theorem 4.6: *If Ω is given as (4.2) and Assumption 4.5 is satisfied, and if a merit function is drawn from the class $\mathcal{M}$, then there exists a small pair $(\alpha, \varepsilon) = (\alpha^*, \varepsilon^*)$ such that the NewRW-Heu method with this pair guarantees to find a sparsest solution of the linear system when either $\vartheta = 1$ or $\vartheta > 1$ and $|u - v| \leq (u + v)/\gamma(w)$, where (u, v, w) is a solution to (3.5).* While we focus our analysis on (4.1) for simplicity in this section, it is possible to extend this analysis to the case where Ω^k is updated iteratively. In fact, consider the algorithm using the scheme (3.8). At the current iterates $(y^k, w^k, s^k, u^k, v^k, x^k)$, the set Ω^{k-1} is changed to Ω^k according to (3.8), and for this given set $\Omega = \Omega^k$, the problem (3.6) will be solved by selecting $\alpha_k < \alpha_{k-1}$. We may choose α_k sufficiently small. This is equivalent to starting from α_{k-1} and applying the simple updating scheme $\alpha_+ = \tau\alpha_-$ repeatedly for a finite number of times, where $\tau \in (0, 1)$ is a fixed constant. Thus by replacing the role of Ω by Ω^k and a suitable design of the algorithm, it is possible to extend the convergence analysis in this section (under some modifications) to the case where Ω^k is updated by (3.8). The performance of Algorithm 3.1 using (3.8) is also demonstrated in the next section.

5. Numerical simulations. In this section, we demonstrate the performance of the algorithm proposed in this paper. Algorithm 3.1 is a general algorithmic framework which admits a large freedom in the choices of $(\alpha, \varepsilon, \Omega, \Phi_\varepsilon(s))$. For simplicity, we fix ε and update α and Ω iteratively according to certain simple schemes. The updating scheme for Ω is motivated by Lemma 4.7. Thus we consider the following specific version of Algorithm 3.1, which is called the NewRW algorithm in this section.

Algorithm 5.1 (NewRW). Given $\alpha_0 \in (0, 1), \tau \in (0, 1), \varepsilon \in (0, 1), 0 < \alpha^* \ll \alpha_0$ and $\vartheta \geq 1$; perform the following steps:

Step 0. (Initialization) Let x^0 and γ^0 be a solution and the optimal value of ℓ_1-minimization, respectively. Let $\Gamma^0 \geq \vartheta \max\{1, 1/\gamma^0\}$ be given, and let

$$\Omega^0 = \{w \in R^n_+ : \ |x^0|^T w \leq \vartheta, \ w \leq \Gamma^0 e\}. \tag{5.1}$$

Step 1. If $\alpha_k \leq \alpha^*$, stop; otherwise, solve the convex optimization problem (3.6) to obtain the vector $(w^{k+1}, y^{k+1}, s^{k+1}, u^{k+1}, v^{k+1})$.

Step 2. Set $W^{k+1} = \mathrm{diag}(w^{k+1})$. Solve the weighted ℓ_1-minimization problem (3.7) to obtain (x^{k+1}, γ^{k+1}).

Step 3. Set $\alpha_{k+1} = \tau\alpha_k$ and $\Gamma^{k+1} \geq \vartheta \max\{1, \|w^{k+1}\|_\infty/\gamma^{k+1}\}$ and let

$$\Omega^{k+1} = \{w \in R^n_+ : \ |x^{k+1}|^T w \leq \vartheta, \ w \leq \Gamma^{k+1} e\}. \tag{5.2}$$

Replace k by $k + 1$ and return to *Step 1.*

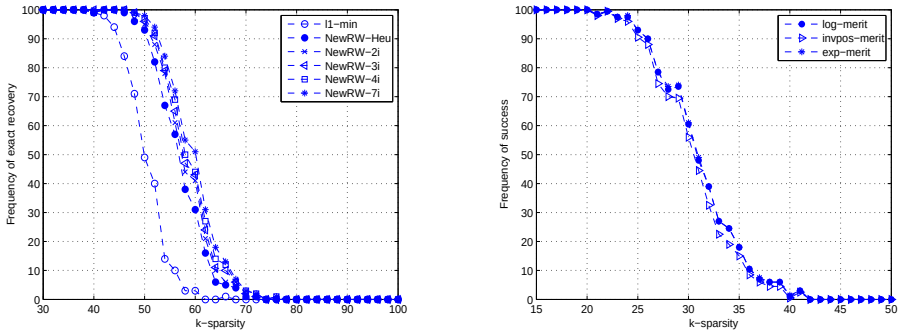
In our experiments, the parameters in Algorithm 5.1 are specified as follows: $\alpha^0 = 10^{-8}, \tau = 0.1, \varepsilon = 10^{-15}$, and $\vartheta = 10^3$. Γ^0 and Γ^{k+1} are given by $\Gamma^0 = \vartheta \max\{1, 1/\gamma^0\} + 1$ and $\Gamma^{k+1} = \vartheta \max\{1, \|w^{k+1}\|_\infty/\gamma^{k+1}\} + 1$. We may perform the

algorithm any prescribed number of iterations by choosing a sufficiently small tolerance α^*. (For instance, under the above choices of parameters the algorithm will be executed a total of 5 iterations when $\alpha^* = 10^{-13}$ is used.) CVX, a package for solving convex programs [33], is employed to solve the problem (3.6) and the weighted ℓ_1-problem in our implementation. The experiments have been carried out by realizing the entries of $A \in R^{m\times n}$ and the nonzero entries of the sparse vector x^* from the standard normal distribution, unless otherwise stated. For each realized pair (A, x^*), we set $b = Ax^*$ and then apply the algorithms to the system $Ax = b$ to test their performance for reconstructing x^*. When the sparsity level of x^* is low, x^* is often the unique sparsest solution to linear systems, so the recovery performance of an algorithm may reflect its capability of solving ℓ_0-minimization problems. In our experiments, we say that x^* is exactly recovered by an algorithm if the solution x found by the algorithm satisfies the recovery criterion

$$\mathrm{RC} = \|x - x^*\|_2^2/\|x^*\|_2^2 \le 10^{-6}.$$

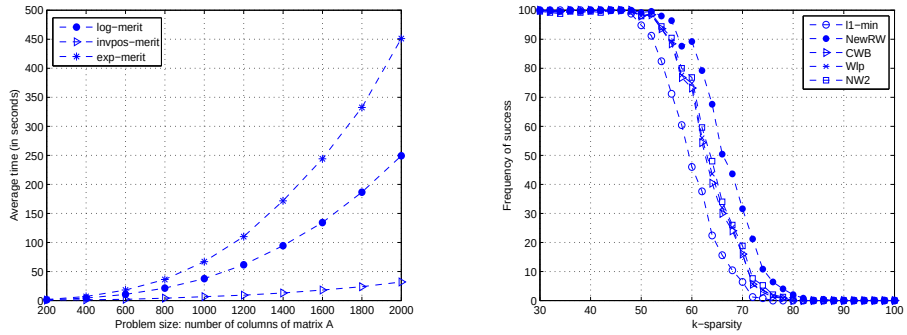
As pointed out in Remark 4.9, performing one iteration of Algorithm 3.1 or 5.1 leads to the following heuristic method: *Given small parameters* (α, ε) *and a set* Ω, *solve the convex problem (3.6) to generate a weight, and solve the weighted* ℓ_1*-problem with this weight to obtain a sparse solution.* This is referred to as the NewRW-Heu method in this section. Our first experiment has been carried out to show the recovery performance of the NewRW-Heu and how the performance of the NewRW is improved as more than one iteration is performed. Specifically, we compare the performance of the NewRW-Heu and the NewRW algorithm with a total of 2, 3, 4 and 7 iterations, respectively (these algorithms are referred to as the NewRW-2i, -3i, -4i and -7i, respectively). For this experiment, the sparsity level k of the vector $x^* \in R^{1000}$ ranges from 30 to 100 according to $k = 30 + 2l$ for $l = 0, 1, ..., 35$, and for each sparsity level we run 200 trials of the linear systems with random matrices of size 200×1000. The results are given in Fig. 5.1(i). It can be seen that the NewRW-Heu, i.e., a single iteration of Algorithm 5.1, remarkably outperforms the standard ℓ_1-minimization, and performing every further iteration of the NewRW may improve the recovery performance of the algorithm. Such an improvement can be remarkable during the first few iterations, but less remarkable as the number of iterations continues to increase. (This phenomenon has also been observed for other reweighted ℓ_1-methods in the literature.) The simulations indicate that for most sparsity problems, it is sufficient to run a few (e.g., 2 to 5) iterations of the NewRW.

We have also tested the performance of the NewRW with three different merit functions (3.10)–(3.12), which are called the exp-, invpos-, and log-merit functions, respectively. The frequencies of exact recovery are included in Fig. 5.1 (ii). It can be seen that the recovery performance of the NewRW method is insensitive to the choice between the merit functions (3.10)–(3.12). However, in terms of the time required for solving the convex problem (3.6), the choice of merit functions does matter when the dimension of the problem is high. Let the size of $A \in R^{m\times n}$ vary from $(m, n) = (40, 200)$ to $(400, 2000)$ according to $(m, n) = (40k, 200k),\ k = 1, ..., 10$. For each of these dimensions, we generate 100 pairs of (A, x^*) in random where the sparsity level of x^* is 20. The average time required for solving the convex problem (3.6) with different merit functions and dimensions is shown in Fig. 5.2(i), from which we see that as the dimension increases, the time required for solving (3.6) with the invpos-merit function is remarkably less than the time for solving the problem with the log-merit function and the exp-merit function. Thus the invpos-merit function is used as a default in our NewRW algorithm.



(i) (ii)

FIG. 5.1. *(i) Exact recovery performance of the NewRW with a fixed number of iterations (1,2,3,4 and 7 iterations). The performance of ℓ_1-minimization is also included for comparison. The experiment was carried out for the linear systems with random matrices in $R^{200\times 1000}$ and 200 attempts were made for each sparsity level $k = 30, 32, 34, ..., 100$. (ii) Exact recovery performance of the NewRW method with different merit functions. The algorithms were performed a total of 5 iterations, and the experiment was carried out for the problems with random matrices in $R^{100\times 500}$ and 200 attempts were made for each sparsity level $k = 15, ..., 50$.*



(i) (ii)

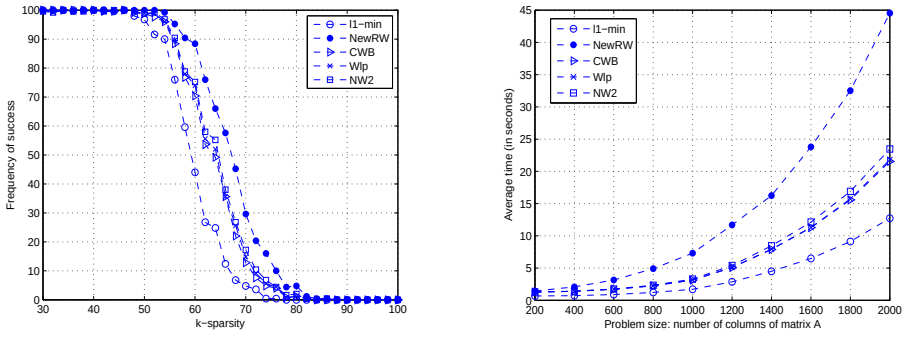
FIG. 5.2. *(i) Comparison of the average time required for solving problem (3.6) with different problem dimensions and with different merit functions. The average is taken over 100 trials for every specified dimension ranged from (m, n)=(40, 200) to (400, 2000). (ii) Exact recovery performance of the NewRW, ℓ_1-minimization, CWB, Wlp, and NW2 algorithms. Experiments were carried out for the systems with dimensions $(m, n) = (200, 800)$, and 300 trials were run for each sparsity level $k = 30, 32, ..., 100$. The sparse vectors x^* were drawn from a normal distribution.*

We now compare the performance of the NewRW method and several existing reweighted ℓ_1-methods. The first reweighted ℓ_1-algorithm using the weight $w^{k+1} = 1/(|x^k| + \rho)$, where $\rho > 0$ is a fixed parameter, was proposed in [12]. This method, referred to as the CWB, has been widely used in the literature. Another reweighted ℓ_1-algorithm being widely used is proposed in [30] and referred to as the Wlp; it uses the weight $w^{k+1} = 1/(|x^k| + \rho)^p$, where $p \in (0, 1)$ is a given parameter. Also, the reweighted ℓ_1-algorithm NW2 presented in [59] is also included here for comparison.

This method uses the weight

$$w^{k+1} = \frac{q + (|x_i^k| + \rho)^{1-q}}{(|x_i^k| + \rho)^{1-q}\left[|x_i^k| + \rho + (|x_i^k| + \rho)^q\right]^{1-p}},$$

where (p, q) are given parameters. In our experiments, we set the parameters $\rho = 10^{-3}$ and $p = q = 0.05$, and the standard ℓ_1-minimizer is used as the initial point, and all algorithms are performed a total of 5 iterations. We run 300 tries of linear systems with $A \in R^{200\times 800}$. As the nonzero entries of the sparse vector $x^* \in R^{800}$ are drawn from the standard normal distribution, the exact recovery performance of these algorithms is shown in Fig. 5.2(ii). When the sparse vectors $x^* \in R^{800}$ are drawn from the uniform distribution over [0,1], the performance of algorithms is demonstrated in Fig. 5.3(i), which seems quite similar to the result in Fig. 5.2(ii). Compared with existing CWB, Wlp and NW2 methods, it can be seen from both figures that NewRW is one of the very efficient algorithms in the family of reweighted ℓ_1-methods in terms of sparsity recovery performance. We believe that the performance of the algorithms still has room for improvement in terms of the updating schemes for $(\alpha, \varepsilon, \Omega)$ and the design of Algorithm 3.1. This is a worthwhile future study.



(i) (ii)

FIG. 5.3. *(i) Exact recovery performance of the NewRW, ℓ_1-minimization, CWB, Wlp, and the NW2 algorithms. Experiments were carried out for the systems with $(m, n) = (200, 800)$, and 300 trials were run for each sparsity level $k = 30, 32, ..., 100$. Sparse vectors were drawn from a uniform distribution. (ii) Comparison of runtime between the NewRW and a few reweighted ℓ_1-methods. The size of random matrices $A \in R^{m\times n}$ varies from $(m, n) = (40, 200)$ to $(400, 2000)$ according to $(m, n) = (40k, 200k)$, $k = 1, ..., 10$. The sparsity level of x^* is 20. The average runtime for each given dimension is taken over 100 trials.*

Moreover, the average time required for performing one iteration of NewRW (with invpos-merit function) and a few existing methods is given in Fig. 5.3(ii) which indicates that the average runtime for NewRW with invpos-merit is reasonably higher than CWB, Wlp and NW2. Note that a convex problem (i.e., (3.6)) and a reweighted ℓ_1-problem are solved within one iteration of Algorithm 3.1. It can be observed from Fig. 5.3(ii) that the runtime of Algorithm 3.1 is roughly twice as much of those purely linear-program-based reweighted ℓ_1-methods.

6. Conclusions. The relation between the sparsest solution of a linear system and weighted ℓ_1-minimization has been clarified in section 2 of this paper. Through the linear programming theory, we have shown that seeking the sparsest solution of a linear system can be achieved by locating the densest slack variable of the dual

problem of weighted ℓ_1-minimization among all possible choices of weights. As a result, ℓ_0-minimization can be transformed to ℓ_0-maximization with certain bilevel constraints. Based on this observation, we have developed a new reweighted ℓ_1-algorithm, going beyond the framework of existing sparsity-seeking methods. We have shown that the proposed algorithm converges to the sparsest solutions of linear systems under some assumptions that do not require a linear system to admit a unique sparsest solution. Assumption 4.5 is used for the efficiency analysis of sparsity-seeking methods for the first time. The numerical simulations indicate that the proposed algorithms outperform the standard ℓ_1-minimization method and are comparable to some existing reweighted ℓ_1-algorithms in solving ℓ_0-minimization problems.

REFERENCES

[1] M.S. Asif and J. Romberg, Fast and accurate algorithms for reweighted ℓ_1-norm minimization, *IEEE Trans. Signal Process.*, 61 (2013), pp. 5905–5916.

[2] M.S. Asif and J. Romberg, Sparse recovery of streaming signals using ℓ_1-homotopy, ArXiv, June 2013.

[3] B. Babadi, D. Ba, P. Purdon and E. Brown, Convergence and stability of a class of iteratively reweighted least squares algorithms for sparse signal recovery in the presence of noise, Technical Report, MIT, 2013.

[4] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorihm for linear inverse problems, *SIAM J. Imaging Sci.*, 2 (2009), pp. 183–202.

[5] A. Beurling, Sur les intégrales de Fourier absolument convergentes et leur application à une transformation fonctionelle, in *Proc. Scandinavian Math. Comgress*, Helsinki, Finland, 1938.

[6] T. Blumensath and M. Davies, Gradient pursuit, *IEEE Trans. Signal Process.*, 56 (2008), pp. 2370–2382.

[7] T. Blumensath, M. Davies and G. Rilling, *Greedy algorithms for compressed sensing*, in Compressed Sensing: Theory and Applications (Y. Eldar and G. Kutyniok Eds.), Cambridge University Press, 2012.

[8] A.M. Bruckstein, D. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Rev.*, 51 (2009), pp. 34–81.

[9] E. Candès, J. Romberg and T. Tao, Robust uncertainty principles: Exact signal reconstruction from higly incomplete frequency information, *IEEE Trans. Inform. Theory*, 52 (2006), pp. 489–509.

[10] E. Candès, J. Romberg and T. Tao, Stable signal recovery from incomplete and inacurate measurements, *Comm. Pure Appl. Math.*, 59 (2006), pp. 1207–1223.

[11] E. Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51 (2005), pp. 4203–4215.

[12] E. Candès, M. Wakin and S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.*, 14 (2008), pp. 877–905.

[13] C. Carathéodory, Über den Variabilitätsbereich der Koeffizienten von Potenzreihen die gegebene Werte nicht annehmen, *Math. Ann.*, 64 (1907), pp. 95–115.

[14] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Proc. Lett.*, 14 (2007), pp. 707–710.

[15] R. Chartrand and W. Yin, Iteratively reweighted algorithms for compressive sensing, IEEE International conference on Acoustics, Speech and Signal Processing (ICASSP), 2008, pp. 3869–3872.

[16] S. Chen, D. Donoho and M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, 20 (1998), pp. 33–61.

[17] X. Chen and W. Zhou, Convergence of the reweighted ℓ_1 minimization algorithms for ℓ_2-ℓ_p minimization, *Comput. Optim. Appl.*, 59 (2014), pp. 47-61.

[18] A. Cohen, W. Dahmen and R. Devore, Compressed sensing and best k-term aproximation, *J. Amer. Math. Soc.*, 22 (2009), pp. 211–231.

[19] W. Dai and O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, *IEEE Trans. Inform. Theory*, 55 (2009), pp. 2230–2249.

[20] I. Daubechies, M. Defrise and C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, 57 (2004), pp. 1413–1457.

[21] I. Daubechies, R. DeVore, M. Fornasier and C.S. Güntürk, Iteratively reweighted least squares

minimization for sparse recovery, *Comm. Pure. Appl. Math.*, 63 (2010), pp. 1–38.
[22] R.A. DeVore and V.N. Templyakov, Some remarks on greedy algorithms, *Adv. Comput. Math.*, 5 (1996), pp. 173–187.
[23] D. Donoho, Denoising by soft-thresholding, *IEEE Trans. Inform. Theory*, 41 (1995), pp. 613–627.
[24] D. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, 52 (2006), pp. 1289–1306.
[25] A. Donoho, I. Drori, Y. Tsaig and J. Starck, Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit, Technical Report, Stanford University, 2006.
[26] D. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, *Proc. Natl. Acad. Sci.*, 100 (2003), pp. 2197–2202.
[27] D. Donoho and X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inform. Theory,* 47 (2001), pp. 2845–2862.
[28] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, 2010.
[29] M. Figueiredo, J. Bioucas-Dias and R. Nowak, Majorization-minimization algorithms for wavelet-based image restoration, *IEEE Trans. Image Process.*, 16 (2007), pp. 2980–2991.
[30] S. Foucart and M. Lai, Sparsest solutions of undertermined linear systems via ℓ_p-minimization for $0 < q \leq 1$, *Appl. Comput. Harmon. Anal.*, 26 (2009), pp. 395–407.
[31] J.J. Fuchs, On sparse representations in arbitrary redundant bases, *IEEE Trans. Inform. Theory,* 50 (2004), pp.1341–1344.
[32] I. Gorodnitsky, J. George and B. Rao, Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm, *Electroen. Clin. Neuro.*, 95 (1995), pp. 231–251.
[33] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming*, Version 1.21, April 2011.
[34] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, NY, 2001.
[35] P. Holland and R. Welsch, Robust regression using iteratively reweighted least-squares, *Comm. Statist. Theory Methods*, A6 (1997), pp. 813–827.
[36] X. Huang, Y. Liu, S. Shi, S. Van Huffel and J. Suykens, Two-level ℓ_1 minimization for compressed sensing, KU Leuven, Belgium, 2013.
[37] M. Khajehnejad, W. Xu, A. Avestimehr and B. Hassibi, Improved sparse recovery thresholds with two-step reweighted ℓ_1 mininimization, *Proc. Int. Symp. on Inform. Theory* (ISIT), Austin, Texas, June, 2010.
[38] M. Khajehnejad, W. Xu, A. Avestimehr and B. Hassibi, Weighted ℓ_1 mininimization for sparse recovery with prior information, ArXiv 2009.
[39] M. Lai and J. Wang, An unconstrainted ℓ_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems, *SIAM J. Optim.*, 21 (2010), pp. 82–101.
[40] Z. Lu, Iterative reweighted minimization methods for ℓ_p regularized unconstrainted nonlinear programming, *Math. Program. Ser. A*, 147 (2014), pp. 277–307.
[41] D. Malioutov and A. Aravkin, Iterative log thresholding, T.J. Watson IBM Research Center, ArXiv, Dec. 2013.
[42] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1999.
[43] S. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.*, 41 (1993), pp. 3397–3415.
[44] O.L. Mangasarian, *Machine learning via polydedral concave minimization*, in Applied Mathematics and Parallel Computing-Festschrift for Klaus Ritter (H. Fischer, B. Riedmueller and S. Schaeffler eds.), Springer, Heidelberg, 1996, pp. 175–188.
[45] O.L. Mangasarian and R. R. Meyer, Nonlinear perturbation of linear programs, *SIAM J. Control & Optim.*, 17 (1979), pp. 745–752.
[46] D. Needell, Noisy signal recovery via iterative reweighted ℓ_1-minimization, In *Proceedings of the 43rd Asilomar conference on Signals, Systems and Computers*, Asilomar'09, 2009, pp. 113–117.
[47] D. Needell and J.A. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples, *Appl. Comput. Harmonic Anal.*, 26 (2008), pp. 301–321.
[48] D. Needell and R. Vershynin, Signal recovery from imcomplete and inaccurate measurements via regularized orthorgonal matching pursuit, *IEEE J. Sel. Topics Sig. Proc.,* 4 (2010), pp. 310–316.
[49] W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*, Van Nostrand Reinhold, 1993.
[50] O. Taheri and S. Vorobyov, Reweighted ℓ_1-norm penalized LMS for sparse channel esitimation and its analysis, arXiv, January 2014.

[51] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. Royal Statist. Soc B*, 58 (1996), pp. 267–288.
[52] J.A. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory*, 50 (2004), pp. 2231–2242.
[53] J.A. Tropp, Just Relax: Convex programming methods for indentifying sparse signals in noise, *IEEE Trans. Inform. Theory*, 52 (2006), pp. 1030–1051.
[54] J.A. Tropp and S.J. Wright, Computational methods for sparse solution of linear inverse problems, in *Proc. of the IEEE,* 98 (2010), pp. 948–958.
[55] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, NY, 1999.
[56] D. Wipf and B. Rao, Sparse Bayesian learning for basis selection, *IEEE Trans. Signal Process.*, 52 (2004), pp. 2153–2164.
[57] D. Wipf and S. Nagarajan, Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions, *IEEE J. Sel. Topics Signal Process.*, 4 (2010), pp. 317–329.
[58] Y.B. Zhao, RSP-based analysis for sparsest and least ℓ_1-norm solutions to underdetermined linear systems, *IEEE Trans. Signal Process.*, 61 (2013), no. 22, pp. 5777–5788.
[59] Y.B. Zhao and D. Li, Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems, *SIAM J. Optim.*, 22 (2012), pp. 1065–1088.